

1 Unified estimator for excess journey time under heterogeneous passenger incidence behavior using 2 smartcard data

3
4 Jinhua Zhao ^{a*}, Michael Frumin ^b, Nigel Wilson ^c, Zhan Zhao ^d

5
6 ^a Department of Civil Engineering, University of British Columbia, Canada

7 * Corresponding Author:

8 Email: jinhua.zhao@ubc.ca; Tel: +1-604-822-2196; Fax: +1-604-822-6901

9 2007 – 6250 Applied Science Lane Vancouver, B.C. V6T 1Z4, Canada.

10 ^b Metropolitan Transportation Authority Bus Customer Information Systems, USA

11 2 Broadway, 27th Floor, New York, NY 10004

12 Tel: 646-252-1117; Email: mfrumin@mtahq.org

13 ^c Department of Civil and Environmental Engineering, MIT, USA

14 1-238, 77 Massachusetts Avenue, Cambridge, MA 02139

15 Tel: (617) 253-5046; (612) 626-1341; nhmw@mit.edu

16 ^d Department of Civil Engineering, University of British Columbia, Canada

17 2002 – 6250 Applied Science Lane Vancouver, B.C. V6T 1Z4, Canada.

18 Email: zhanzhaowf@hotmail.com

19 20 21 **Abstract**

22 Excess journey time (EJT), the difference between actual passenger journey times and journey times
23 implied by the published timetable, strikes a useful balance between the passenger's and operator's
24 perspectives of public transport service quality. Using smartcard data, this paper tried to characterize
25 transit service quality with EJT under heterogeneous incidence behavior (arrival at boarding stations). A
26 rigorous framework was established for analyzing EJT, in particular for reasoning about passenger'
27 journey time standards as implied by varying incidence behaviors. It was found that although the wrong
28 assumption about passenger incidence behavior and journey time standards could result in a biased
29 estimate of EJT at the individual passenger journey level, the paper proposed a unified estimator of EJT,
30 which is unbiased at the aggregate level regardless of the passenger incidence behavior (random incidence,
31 scheduled incidence, or a mixture of both). A case study based on the London Overground network (with
32 a tap-in-and-tap-out smartcard system) was conducted to demonstrate the applicability of the proposed
33 method. EJT was estimated using the smartcard (Oyster) data at various levels of spatial and temporal
34 aggregations in order to measure and evaluate the service quality. Aggregate EJT was found to vary
35 substantially across the different London Overground lines and across time periods of weekday service.
36 The North London Line in the AM Peak in the westbound direction had the worst service quality in terms
37 of EJT.

38
39 **Keywords:** excess journey time; service quality; passenger incidence behavior; smartcard data; London
40 Overground

41 42 **Highlights**

- 43 • A rigorous framework was established for analyzing excess journey time (EJT), in particular for
44 reasoning about passenger' journey time standards as implied by varying incidence behaviors.
 - 45 • A unified estimator for aggregate EJT was proposed, which was unbiased at the aggregate level
46 regardless of actual passenger incidence behavior.
 - 47 • The proposed estimator was applied to the London Overground network using Oyster smartcard data.
 - 48 • EJT was calculated at various levels of spatial and temporal aggregations and significant variations
49 between lines and time periods were observed in the London Overground
- 50

1. Introduction

The performance of public transport service can be viewed from different angles. The Transit Capacity and Quality of Service Manual defines *service quality* as “the overall measured or perceived performance of [public transport] service from the passenger's point of view”(Kittelsohn & Associates, 2003). It also points out that the public transport agency or operator often perceives system performance from a different perspective, one more concerned with the quality of the operations than of the service as experienced by passengers. It defines *service delivery* in terms of how well “an agency deliver[s] the service it promises on a day-to-day basis.” In this paper, service quality will refer to the passenger's perspective on system performance while service delivery will refer to the operator's perspective.

With the introduction of automatic fare collection (AFC) systems and the data they produce about individual passenger journeys, it is now possible to measure certain aspects of service quality directly. Some AFC systems (e.g. Oyster system in London) control entry to and exit from the public transport network. In this case, actual passenger journey time through the network can be measured as the difference between the timestamps of the exit and entry transactions.

Direct and automatic observation of passenger journey times creates many opportunities for measuring service quality. One particular measure explored in this paper is excess journey time (EJT). At the level of a single journey, EJT is the difference between actual journey time and some pre-defined journey time standard, however that standard is defined. A positive value indicates that the journey took longer than the standard allows; a negative value indicates that it was shorter. Although actual passenger journey time can be directly captured by AFC systems, the standard cannot. The journey time standard depends on how passengers plan their arrival to public transport services. Because passengers can also arrive at certain location via public transport services, a lexical convention is established to avoid ambiguity of exposition. *Passenger incidence* is defined here as the act or event of being incident to a public transport service with intent to use that service (Frumin and Zhao, 2012). It is very difficult to track the incidence behavior of individual passengers. Certain simplifying assumptions on passenger incidence behavior are made to set the appropriate standard for passenger journey time. However, in reality, passenger incidence behavior is often more heterogeneous than any of the assumptions indicates. This poses challenges in how realistic EJT can be, and how transit agencies can implement it in their practice.

The objective of this paper is to use smartcard data to measure public transport service quality as experienced by passengers in terms of EJT, and investigate the sensitivity of the estimator to the two types of passenger incidence behaviors (random incidence and scheduled incidence). The research presented in this paper has been developed commensurate with the analytical needs of the managers of the rapidly growing, largely circumferential London Overground network in London, England. The methods developed in this paper are applied to smartcard data from this railway. A case study is conducted to demonstrate the use of these methods to measure and evaluate the service quality on the core portion of the Overground network. While the work in this paper is motivated primarily by problems facing the managers and passengers of London Overground, the methods it develops should generalize well to other transit systems with off-board entry and exit control (like some BRT systems) where similar automatic data are available.

The rest of the paper is organized in this way. Section 2 reviews some of the literature of public transport service delivery and service quality measurement, including EJT. Section 3 discusses the implications of incidence behavior for establishing EJT standards, and proposes an analytical framework with which to analyze EJT under different incidence behaviors. Then a rigorous probabilistic analysis is used to prove, under the proposed framework, that a single means for measuring aggregate EJT appropriately can accommodate a range of incidence behaviors. Section 4 applies the method developed here to the London Overground network. Section 5 draws conclusions and discusses some important considerations for applying EJT.

2. Literature review

The literature on measurement of public transport performance, including service delivery and service quality, is rich. Interest in the subject has renewed since the introduction of systems for automatically monitoring various aspects of public transport operations and, most recently, public transport passenger journeys. The primary client of public transport performance measurements are the managers and planners of the transport networks themselves. Ideally, they should be motivated to improve the service quality as experienced by their primary customers, the passengers. However, the levers over which they have the most tangible understanding and direct control are planning and operational service delivery. Consequently, it is proposed that measures of public transport performance should find a balance between the passenger's and operator's perspectives. They should strive for fidelity to the passenger experience, but not so far that they are not useful or interpretable by operators.

2.1. Reliability

Before discussing the literature on specific measures of service delivery and service quality (i.e. from the operator's and passenger's perspectives) this section discusses first the notion of *reliability*, defined as "the invariability of service attributes which influence the decisions of travelers and transportation providers." (Abkowitz *et al.*, 1978) Under this definition, the discussion of reliability is quite naturally subsumed by discussions of service delivery and quality if and when they consider higher-order moments of the attributes perceived by operators and by passengers, respectively. Consequently, much of what has been said about reliability applies to both service delivery and quality, and thus applies to the balance of this paper. Understanding reliability as a proxy for overall performance, including service delivery and service quality, Abkowitz *et al.* (1978) note that measuring performance from the operator's and passenger's perspectives should help public transport providers to: (i) identify and understand performance problems; (ii) identify and measure actual improvements in performance; (iii) relate such improvements to particular strategies; (iv) modify strategies, methods, designs to obtain greater performance improvements." In the context of this paper, this description is useful in that it establishes the measurement of service delivery and service quality as elements of an iterative analytical management and planning process.

2.2. Service delivery measurement and the operator's perspective

In terms of service delivery, Furth *et al.* (2006) describe two classes of service delivery measures that could be developed from automatic data sources: those measuring adherence to timetables and those measuring adherence to headways.

2.2.1. Timetable-based measures

Timetable-based measures are often based on observations of schedule deviation - the difference, for a given service, between the scheduled and actual time of arriving, passing, or departing a given time point. The most popular measure of timetable adherence is on-time performance (OTP), the fraction of services with schedule deviation within some threshold (Kittelson & Associates, 2003). Under the name of Public Performance Measure (PPM), this is the current measure of performance on the London Overground and all other National Rail services in the UK, with a train considered "on time" if it is less than 5 minutes late at the destination terminal.

Henderson *et al.* (1990) and Henderson, Adkins and Kwong (1991) offer a number of criticisms of OTP, primarily for its lack of passenger orientation. Among these criticisms are (i) OTP measures performance at terminals, which for many networks are remote from the locations to which most passengers are bound, (ii) OTP typically counts as late services which have missed part of their trip or skipped stops, even if these adjustments don't affect many passengers, (iii) passenger waiting times are not accurately reflected, (iv) focusing on OTP can incentivize dispatch actions that favor schedule

1 adherence over regular headways and can make passengers worse off, and (v) OTP, while offering a
2 probabilistic measure, does not represent the odds of on-time arrival realistically.

3 A related measure is terminal-to-terminal running time. Statistics of the distribution of running time
4 indicate the reliability of a service and are important inputs into the scheduling process. When running
5 times are too short, some vehicles will not be in place to make subsequent trips; when they are too long,
6 resources are not used efficiently and terminals may be congested (Furth *et al.*, 2006; Furth and Muller,
7 2007).

8 Another common timetable-based service delivery measure is en-route schedule adherence (ESA),
9 which can be defined as the fraction of services with schedule deviation within some threshold at a given
10 set of time points. This is similar to OTP, but applied at multiple points along a line. The distribution of
11 schedule deviation and segment (i.e. point-to-point) running times can also be studied (Furth *et al.*, 2006;
12 Hammerle *et al.*, 2005).

14 2.2.2. Headway-based measures

15 Some public transport network or lines publish service headways but not timetables. In some cases,
16 particularly for higher frequency services, it is assumed that passengers do not use the timetable even if it
17 is available. Kittelson & Associates (2003) recommend measuring the mean observed headway and the
18 coefficient of variation with respect to the mean scheduled headway.

19 Henderson, Kwong and Heba (1991) propose two measures of headway regularity, one based on Gini's
20 ratio and the other based on the coefficient of variation, that have the benefit of being normalized on a
21 zero to one scale for comparison across services with different mean headways. These measures are all
22 unitless, and thus hard to interpret in physical terms relevant to operators or passengers (Furth *et al.*,
23 2006).

24 Reddy *et al.*(2009) and Hammerle *et al.*(2005) define headway regularity in terms of the fraction of
25 observed headways that are within some absolute or relative deviation from the scheduled headway.
26 These have the benefit of being easy to interpret by operators, but still fail to translate easily into
27 passenger terms (Furth *et al.*, 2006).

28 The adoption of headway-based measures is motivated by the effect of headways on passenger waiting
29 times, and so is a real step towards representing the passenger's perspective. Nevertheless, they are still an
30 indirect proxy for the passenger experience, since waiting times are related to but not equal to headways.
31 Moreover, headway-based measures do not account for the entire duration of passenger journeys, which
32 are important. Finally, headway at a given location depends on which services one is willing to board at
33 that location (e.g. for trunk-and-branch services), which depends on where one is headed (Frumin and
34 Zhao, 2012). Headways cannot be accurately measured without considering the passenger's perspective.

36 2.3. Service quality measurement and the passenger's perspective

38 Strictly speaking, service quality is absolute in nature, at least with respect to service delivery. For
39 example, the service quality of a public transport network can be judged on its waiting and travel times.
40 Even when every passenger experiences perfect service delivery - more frequent and faster service is
41 always better. The work described in this section seeks to measure service quality in absolute terms.

42 For randomly incident passengers, Osuna and Newell (1972) describe how mean waiting times can be
43 modeled given observations of actual headways. Friedman (1976) extends this result to model the
44 variance of waiting times. Larson and Odoni (2007) describe how the complete distribution of waiting
45 times of randomly incident passengers can be derived from headway observations.

46 Bates *et al.*(2001) provide an in-depth investigation of how passengers value reliability (expressed as
47 the variability of total journey time) and how it may affect their behavior. Furth and Muller (2006)
48 operationalize some of this analysis by proposing to measure the effect of reliability as additional waiting
49 time costs perceived by passengers. Their analysis is based on the finding of Bates *et al.* (2001) that
50 passengers adjust their incidence behaviour based on knowledge of schedule and headway adherence and
51 reliability. For short headway services, they propose to use headway observations to measure the

1 “potential waiting time” as the difference between the “budgeted” 95th percentile waiting time and the
 2 mean waiting time. It represents an additional penalty that passengers pay for the unreliability of the
 3 service headways, albeit a penalty paid in many cases by arriving early at their destination.

4 Chan(2007) and Wilson *et al.* (2008) extend the potential waiting time concept to the entire journey.
 5 They use data from the Oyster smartcard ticketing system to measure (rather than model) the journey
 6 times of London Underground passengers. They estimate the distribution of end-to-end journey times for
 7 each origin-destination (OD) station pair and find the “reliability buffer time” (RBT) as the difference
 8 between the 95th and 50th (median) percentiles. This metric is aggregated from the OD pair to the line or
 9 network level by means of an OD flow-weighted average.

10 Uniman (2009) notes that some “irreducible” amount of variability in passenger journey times is to be
 11 expected because of randomness in waiting times, variation in walking speeds, and normal but acceptable
 12 variability in service outcomes. Uniman proposes to divide observation periods into two classes of
 13 reliability levels – “recurrent” and “incident-related.” Passengers experience normal levels of journey
 14 time variability in the former level, and abnormal levels in the latter. Also studying the London
 15 Underground, Uniman makes this classification using a statistical technique that did not consider the
 16 perspectives of the managers of the system under study. Uniman then proposes as a measure of service
 17 quality “excess reliability buffer time” (ERBT) - the difference in RBT for journeys from all observation
 18 periods together and RBT for only those journeys in periods of recurrent reliability. In other words, a
 19 measure of how far the tail of the travel time distribution is extended as a result of abnormal operating
 20 conditions.

21 The measures discussed in this section are developed entirely with reference to actual operating
 22 conditions and passenger journeys, not with reference any service delivery commitments (i.e. the
 23 timetable, and headways and travel times implied therein). No evidence has been found that any of these
 24 measures are regularly used in practice by public transport providers. Such measures may not yet have
 25 been adopted because they do not provide information in terms that operators can easily relate to.

26 2.4. Relative service quality

27
 28
 29 The measures of relative service quality described represent a compromise between the pure operator
 30 and passenger perspectives. They measure service quality not in absolute terms, but rather with respect to
 31 certain standards derived from service delivery commitments.

32 Wilson *et al.* (1992) propose “excess waiting time” (EWT), i.e. the difference between these the actual
 33 passenger waiting time and the expected waiting time that would result from perfect adherence to
 34 schedule. London Transport (1999) extended the EWT concept to the entirety of journeys on the London
 35 Underground, comparing mean actual and schedule values of each component of passenger journeys.
 36 Automatic data from train control systems is used to estimate EWT, as in Wilson *et al.* (1992), under
 37 random incidence model. The EWT estimate is augmented by models for estimating the fraction of
 38 passengers, based on static demand data, who are left behind by overcrowded trains. Automatic train
 39 movement data is also used to estimate excess on-train time, where the scheduled on-train time between
 40 any given pair of stations is as per the timetable. Manual sampling at 27 major stations is combined with
 41 pedestrian flow models to estimate access, egress, and interchange (i.e. walking) time as a function of
 42 pedestrian congestion and availability of escalators and elevators. The scheduled values for pedestrian
 43 movements are determined from manual samples under free-flow conditions. In result, the sum of these
 44 components is referred to as “excess journey time” (EJT), i.e. the difference between the median journey
 45 time and the scheduled journey time.

46 Chan (2007) and Wilson *et al.* (2008) use Oyster journey data to directly estimate (rather than model)
 47 unweighted EJT for individual journeys on the London Underground. They measure actual journey times
 48 directly from Oyster transactions, and derive scheduled journey time from the values in the
 49 Underground's existing EJT measurement system. The results suggest that, given common scheduled
 50 journey time values, unweighted Oyster-based EJT will be more accurate than model-based estimates.
 51 Besides, Buneman (1984) uses schedule-based assignment to estimate passenger on-time performance for

1 the BART railway network in the San Francisco Bay Area. The difference between actual and scheduled
2 arrival time, in the parlance of this section, is an estimate of schedule-based EJT. Buneman does not
3 calculate aggregate EJT, but rather compares EJT to a 5-minute threshold window to estimate passenger
4 OTP. It appears that this measure, perhaps in a modified form, is still used by BART over two decades
5 later.

6 The measures of relative service quality represent a compromise between the pure operator and
7 passenger perspectives. One of these measures, EJT, has found lasting application in large urban railways
8 such as the London Underground and BART networks. It not only presents a compelling alternative to the
9 train on-time performance (OTP) measure currently used by the London Overground, but also measures
10 the actual passenger experience in terms of end-to-end journey time, but reports it with respect to certain
11 service quality standards.

12 All of the measures of relative service quality discussed in this section were developed with the intent
13 of representing the passenger's perspective. However, they all make certain assumptions about passenger
14 incidence behavior. Based on these assumptions they derive the standards against which measured or
15 modeled service quality is compared.

17 *2.5. Heterogeneous passenger incidence behavior*

18
19 With the advent of AFC ticketing systems, actual journey times can now be measured simply and
20 directly as in Wilson *et al.* (2008) and Uniman (2009). One issue that remains unresolved, particularly as
21 EJT is applied to networks with lower service frequencies, is how passenger behavior and expectations
22 relate to the published or unpublished timetable, and thus how the timetable should be used in setting
23 journey time standards.

24 Industry manuals (e.g. Kittelson & Associates, 2003); Furth *et al.*, 2006) typically recommend
25 timetable-based measures for lower frequency services with a headway greater than 10 minutes, where
26 passenger incidence is assumed to be timetable-dependent, and headway-based measures for higher
27 frequency (i.e. shorter headway) services, where passenger incidence is assumed to be random. London
28 Buses, for example, follows this pattern, classifying bus routes as "high frequency" at frequencies of 5 or
29 more buses per hour (a 12 minute or lower headway), and "low frequency" otherwise (Camilletti, 1998;
30 Camilletti, 2003).

31 Most of the relative service quality measures discussed here, including EJT on the London
32 Underground, use the random incidence assumption to derive waiting time standards. The model of
33 Buneman utilizes mixed assumptions about passenger incidence behavior to derive waiting time standards,
34 but he acknowledges that they are arbitrary. These various approaches depend, explicitly or implicitly, on
35 assumptions regarding how passengers' knowledge of the timetable affects their arrival behavior at rail
36 stations and their expectations of waiting and travel time (and distributions thereof).

37 The stated intent of these recommendations and practices is to match journey time standard to the
38 concerns, experiences, and expectations of passengers. The standards against which measured or modeled
39 service quality is compared have been explicitly derived from these simplifying assumptions about
40 passenger incidence behavior. However, passenger incidence behaviors, let alone passenger expectations,
41 are in many cases not so clear cut (Frumin and Zhao, 2012). It is possible to have a mix of timetable-
42 dependent and timetable-independent passengers using the same service at the same time. In cases when
43 behavior is homogeneous across some segments of passengers (e.g. those traveling between a given pair
44 of stations), it still possible to have varying conditions across the network or even at a given station.
45 Trunk-and-branch services, which provide different service frequencies to different passengers at the
46 same departure station, are a prime example. Moreover, incidence behaviors are likely to change over
47 time as a function of changes in relevant attributes of the service (e.g. headway and reliability). Even
48 where the random incidence assumption has historically been justified by a lack of posted timetables (e.g.
49 the London Underground), the reality may be changing as a result of internet and mobile delivery of
50 timetable information.

1 Frumin and Zhao (2012) proposed a method to estimate incidence headway and waiting time by
 2 integrating disaggregate smartcard data with published time tables using schedule-based assignment and
 3 applied it to stations in the entire London Overground to demonstrate its practicality and observe that
 4 incidence behaviour varies across the network and across times of day, reflecting the different headways
 5 and reliability. They classify passenger incidence behaviour into two types – scheduled incidence
 6 passengers (passengers whose incidence is timetable-dependent) and random incidence passengers
 7 (passengers whose incidence behaviour is random and completely independent of scheduled departure
 8 times).

9 This heterogeneity of incidence behavior is a reason that existing measures of service delivery and
 10 (absolute or relative) service quality often fail to appropriately account for the passenger's experience. It
 11 presents a particular problem in measuring EJT, where different assumptions about incidence behavior
 12 could lead to very different journey time standards. This paper proposes and explores a methodology for
 13 estimating aggregate EJT that, it turns out, applies equally well under a range of assumptions regarding
 14 passenger incidence and implied journey time standards.

17 3. EJT-based service quality measurement framework

18
 19 This section proposes an analytical framework, with which EJT estimators are developed at both the
 20 individual level and the aggregate level. These estimators are then compared under different incidence to
 21 show how EJT would change under different passenger incidence assumptions.

23 3.1. Analytical framework and assumption

24
 25 For clarity of exposition, the following lexical convention is adopted. The expectation of a given
 26 quantity refers to the expected value of that quantity in the probabilistic sense. The standard for a given
 27 quantity refers to some individual's supposition of what that quantity should be. Standards can be random
 28 or deterministic. In this discussion, random variables will be shown as capitals, X , known quantities as
 29 lowercase, x , and standards as capitals with tildes, \tilde{X} . The following analysis considers only trips along a
 30 single line without interchanges.

31 For a given passenger, let

32 I = the time that passenger is incident at his or her boarding station,

33 \tilde{W} = the standard for waiting time, also referred to as the scheduled waiting time,

34 \tilde{V} = the standard for in-vehicle travel time, also referred to as the scheduled travel time,

35 \tilde{A} = the standard arrival time at the alighting station, also referred to as the scheduled arrival time,

36 \tilde{J} = the standard for end-to-end journey time from incidence at the boarding station to arrival time at
 37 the alighting station, also referred to as the scheduled journey time,

38 J = the observed or actual journey time,

39 X = the Excess Journey Time (EJT).

40 With these definitions, the following equations establish the intuitive analytical framework:

$$41 \quad \tilde{A} = I + \tilde{W} + \tilde{V} \quad (1)$$

$$42 \quad \tilde{J} = \tilde{A} - I \quad (2)$$

$$43 \quad X = J - \tilde{J} \quad (3)$$

44
 45
 46 Equation (1) says that the arrival time standard is the incidence time plus some standard for waiting
 47 time plus some standard for in-vehicle time. Equation (2) says that the journey time standard is the arrival
 48 time standard less the incidence time. Equation (3) simply formalizes the definition of EJT. Naturally, the
 49 first two equations imply that the journey time standard is the sum of the waiting time standard and the in-
 50 vehicle time standard, i.e. $\tilde{J} = \tilde{W} + \tilde{V}$.

1 Without loss of generality, consider an origin station ("station 1") on a rail line, a randomly selected
 2 passenger traveling from that station to a destination station ("station 2") on the same line, a set of trains
 3 that passenger is willing to board, including a pair of those consecutive trains scheduled to depart from
 4 station 1 towards station 2 with the first train scheduled to depart at time $t = 0$. Figure 1 uses a time-
 5 distance diagram to illustrate the following additional quantities relevant to this analysis.

6 Let

7 d = the scheduled departure time from station 1 of the next train that the passenger in question is
 8 willing to board.

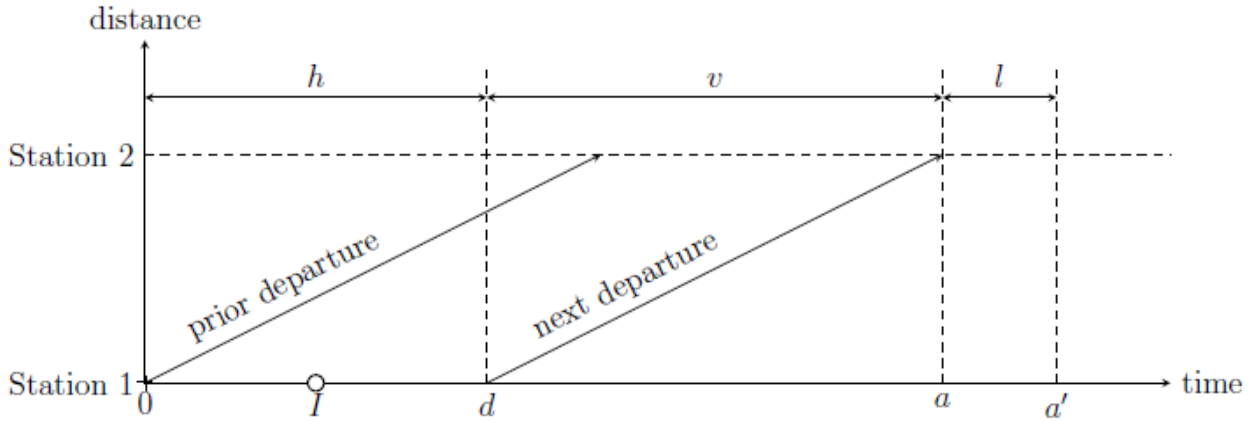
9 h = the scheduled headway between the prior scheduled departure and the next scheduled departure.

10 a = the scheduled arrival time at station 2 of the train departing station 1 at time d .

11 v = the scheduled running time from station 1 to station 2 of the train departing at time d .

12 a' = the actual arrival time at station 2 of the train carrying the selected passenger (whichever train that
 13 may be).

14 l = the difference between the actual arrival time at station 2 of the train carrying the selected
 15 passenger and the scheduled arrival time at station 2 of the train scheduled to depart station 1 at d .



17 **Figure 1.** Example time-distance graph.

18 The following useful relationships are implied by this diagram and related definitions:

19
$$h = d - 0 = d \quad (4)$$

20
$$a = d + v \quad (5)$$

21
$$a' = a + l \quad (6)$$

22
$$J = a' - I \quad (7)$$

23 Let $f_i(i); i \in [0; d]$ be the probability density function for passenger incidence times during the headway
 24 in question. This function is assumed to be continuous, representing a smoothed description of behavior
 25 during the given headway on an average day. It is assumed that all passengers belong to one of the two
 26 behavioral classes of individuals, each with its own method for setting journey time standards. The two
 27 classes will be referred to here as scheduled incidence and random incidence (Frumin and Zhao, 2012).

28 Scheduled incidence passengers are assumed to have knowledge of scheduled departure times and
 29 scheduled running times, which they use both to time their incidence and to set waiting and in-vehicle
 30 time standards. It is assumed that their standard for waiting time is exactly the time between incidence
 31 and the next scheduled departure (i.e. the time they would expect to wait, given their time of incidence,
 32 if they expected the next train to depart as per the timetable), and that their standard for in-vehicle time is as
 33 per the timetable. In the context of the analytical framework, this implies

34
$$\tilde{W}^S = d - I \quad (8)$$

$$\tilde{V}^S = v \quad (9)$$

These results correspond with the simple intuition that if a passenger has knowledge of the timetable, her standards for a given journey depend on her time of incidence only insofar as it determines the next scheduled departure. Her standard for arrival depends only on the timetable for that departure. These equations, along with Equations (1) – (2) and (5) - (7), substituted into Equation (3) yield the similarly intuitive result that

$$X^S = l \quad (10)$$

Consequently, conditioned upon the passenger being incident on the given headway and arriving at time a' , EJT is independent of I and thus is not a random quantity.

Because this class of passengers are assumed to be aware of the schedule, all that is assumed regarding the distribution of their incidence times over a given headway h is that it is not uniform (i.e. completely random). Specifically speaking, a continuous function $f_I^S(i)$ is taken to be a probability density function for the incidence times of scheduled incidence passengers if it meets the following conditions:

$$f_I^S(i) \geq 0, i \in [0, h] \quad (11)$$

$$\exists i \in [0, h]: f^S(i) \neq \frac{1}{h} \quad (12)$$

$$\int_0^h f_I^S(i) di = 1 \quad (13)$$

Random incidence passengers are assumed to have knowledge of scheduled running times and headways but not to have or not use any knowledge of scheduled train departure times. These passengers are assumed to set standards for waiting time based on knowledge of scheduled train headways and to set standards for in-vehicle time based on knowledge of scheduled train running times. Specifically, it is assumed that their standard for waiting time is exactly half the scheduled headway in which they are incident, and that their standard for in-vehicle time is as per the timetable. In the context of the analytical framework, this implies

$$\tilde{W}^R = h/2 \quad (14)$$

$$\tilde{V}^R = v \quad (15)$$

These results correspond with the intuition that if a passenger has no knowledge of specific departure times, his standard for arrival time will depend on his time of incidence, but that his a priori standard for total journey time is independent of his time of incidence. These equations, along with Equations (1) – (2) and (5) - (7), substituted into Equation (3) yield the similarly intuitive result that

$$X^R = l + \frac{h}{2} - I \quad (16)$$

EJT for random incidence passengers is, unlike for scheduled incidence passengers, a random variable, even when conditioned upon being incident in the given headway and arriving at time a' . This result is also intuitive, indicating that the EJT for a given randomly incident passenger depends on luck with respect to how close his time of incidence is to subsequent departures.

For random incidence passengers, conditional upon being incident at a given station during a given scheduled headway, their specific times of incidence are assumed to be uniformly random. In precise terms, for a passenger incident during a given headway h , the classical assumption (e.g. Osuna and Newell, 1972) is made that

$$f_I^R(i) \begin{cases} \frac{1}{h}, i \in [i, h] \\ 0, otherwise \end{cases} \quad (17)$$

3.2. A unified unbiased estimator for aggregate EJT

Section 3.1 reveals that the wrong assumption regarding the class of this passenger biased the estimation of EJT for at the individual passenger journey level. In practice there is no intent to report EJT at the level of an individual passenger. Rather, EJT should be aggregated over many passengers to indicate the performance of all or part of the network in question over a period of time. Of interest is an estimate of the probabilistic expectation (i.e. the mean) of X , $E[X]$ for any given passengers incident on the headway $[0, h]$.

3.2.1. Framework for aggregate EJT

If EJT is to be aggregated over multiple journeys, the analytical framework is insufficient as currently constructed. Different passengers traveling between the same two stations and incident on the same scheduled headway can have different arrival times depending on the actual departure times of trains from the origin station. For example, if, on a given day, trains departed station 1 at times 0 , $h/3$, and h , and some passengers were incident on $[0, h/3]$ and others were incident on $[h/3, h]$, then it is highly unlikely that all passengers incident on $[0; h]$ could have the same arrival time a' .

To account for this, the framework is generalized. It still considers a single randomly selected passenger incident to station 1 on headway $[0, h]$, but rather than a single train arriving at station 2 at time a' , instead consider K discrete trains arriving at station 2. For the k th train, $k \in 1 \dots K$, let

a'_k = the arrival time at station 2 of train k ,

a = as before, the scheduled arrival time at station 2 of the first train scheduled to depart station 1 at the end of the given headway,

l_k = the difference between the arrival time of train k and a ,

Y_k = an indicator random variable, for each passenger, which is 1 if the passenger arrive on train k , 0 otherwise,

α_k = the fraction of all passengers incident at station 1 on $[0, h]$ who arrived at station 2 at a'_k , trivially equal to $E[Y_k]$,

$g_I^k(i)$ be a probability density function, defined over $[0, h]$, describing the distribution of the incidence time of the passengers who were incident on $[0, h]$ and traveled from station 1 to station 2 aboard train k .

It is appropriate to model the arrival times of passengers discretely since train arrivals are a discrete phenomena, at least as compared to passenger incidence. The set of K trains is exhaustive in that it includes all trains used by passengers incident on $[0, h]$ and traveling from station 1 to station 2. This is sufficient to write that

$$\sum_{k=1}^K \alpha_k = 1 \quad (18)$$

$$\sum_{k=1}^K \alpha_k g_I^k(i) = f_I(i) \quad (19)$$

It will also be useful to use the law of total expectation to decompose $E[X]$ as a function of the respective probabilities and conditional expectations of X for passengers arriving on each of the K trains as

$$E[X] = \sum_{k=1}^K \alpha_k E[X|Y_k = 1] \quad (20)$$

3.2.2. Equivalence of random and scheduled incidence assumptions

In Equation (10) it was shown that under the assumption of scheduled incidence, for a given journey incident at station 1 on $[0, h]$ and arriving at station 2 at time a' , EJT is not a random variable but rather equal to 1, independent of time of incidence I . Because the extended framework uses the indicator random

1 variable Y_k , X^S is a random variable. However, conditional upon a given passenger being on train k , EJT
 2 for that passenger is no longer random and is known to be l_k , which implies that

$$3 \quad E[X^S | Y_k = 1] = l_k \quad (21)$$

4
 5
 6 Substituting this into Equation (20) yields, quite intuitively, that under the assumption of scheduled
 7 incidence the estimator for aggregate EJT is

$$8 \quad E[X^S] = \sum_{k=1}^K \alpha_k l_k \quad (22)$$

9
 10
 11 Under the random incidence assumption, it was seen in Equation (16) that EJT for a given journey
 12 does depend on time of incidence. However, conditioned on a specific incidence time $I = i$, X^R is a
 13 deterministic quantity.

14 If it is assumed that the passengers in question are in fact randomly incident, Equations (17) and (19)
 15 can be used to write that

$$16 \quad \frac{1}{h} = \sum_{k=1}^K \alpha_k g_I^k(i), \quad i \in [0, h] \quad (23)$$

17
 18
 19 This along with the fact that the integral of any probability density function over its entire domain
 20 equals unity simplifies the estimator for aggregate EJT under random incidence assumptions to

$$21 \quad E[X^R] = \int_0^h 1/h \left(\frac{h}{2} - i\right) di + \sum_{k=1}^K \alpha_k l_k \quad (24)$$

22
 23
 24 which simplifies further to

$$25 \quad E[X^R] = \sum_{k=1}^K \alpha_k l_k \quad (25)$$

26
 27
 28 which is the same result as found for scheduled incidence in Equation (22). Note that l_k can still be
 29 negative because in this generalized context, some trains departing station 1 on $[0, h]$ may arrive at station
 30 2 before time a .

31 The estimator for aggregate EJT under scheduled incidence assumptions is thus shown to be equal to
 32 the estimator for aggregate EJT under random incidence assumptions if passengers are in fact randomly
 33 incident. This implies that using the scheduled incidence estimator for aggregate EJT is appropriate if all
 34 passengers are scheduled incidence passengers or all passengers are random incidence passengers.

35 3.2.3. Blended passenger incidence behavior

36 In practice, it will often be the case that some passengers are randomly incident while others clearly
 37 make use of the timetable. This would be indicated by a distribution of passenger incidence times over a
 38 given headway that is clearly a superposition of two different incidence distributions, one meeting the
 39 scheduled incidence conditions of Equations (11) - (13), and one meeting the random incidence
 40 conditions of Equation (17).

41 Without loss of generality, assume that some fraction γ of passengers incident on $[0, h]$ are random
 42 incidence passengers, and so $1 - \gamma$ are scheduled incidence passengers. The probability density function
 43 for incidence times of all passengers under blended incidence can then be written as the superposition of
 44 the respective random and scheduled incidence density functions

$$45 \quad f_I^B(i) = \gamma f_I^R(i) + (1 - \gamma) f_I^S(i), \quad i \in [0, h] \quad (26)$$

46
 47
 48
 49 Through a series of derivation, it is found that

$$E[X^B] = \int_0^h 1/h(\frac{h}{2} - i) di + \sum_{k=1}^K \alpha_k l_k \quad (27)$$

which is the same as Equation (24), again simplifying to

$$E[X^B] = \sum_{k=1}^K \alpha_k l_k \quad (28)$$

The estimator for aggregate EJT in the case of multiple trains carrying blended incidence passengers is thus found to be the same as the estimator for aggregate EJT under scheduled incidence assumptions.

A simple example is used to test this finding. Consider a transit service with a constant scheduled headway of 15 min. In reality, each train departs exactly 5 min late (see Figure 2). All passengers are randomly incident. Since both the scheduled and the actual headways are evenly distributed, $E[X^R]$ (or the aggregate EJT for random incidence passengers) should be zero. Based on the derivation above, this result should be the same as $E[X^S]$, the aggregate EJT for scheduled incidence passengers. Assuming passengers incident during (8:00, 8:15] at a constant rate λ (passengers/min) know the specific schedule, all of them will have the same expected arrival time $\tilde{A} = 8:35$. For the 5λ passengers incident during (8:00, 8:05], they will actually arrive at the destination at 8:25, and thus their EJT will be $X_1^S = -10$ min. For the 10λ passengers incident during (8:05, 8:15], they will actually arrive at the destination at 8:40, and thus their EJT will be $X_2^S = 5$ min. Therefore, the aggregate EJT under the scheduled incidence assumption will be $E(X^S) = 5\lambda * X_1^S + 10\lambda * X_2^S = 0$, which is the same with the $E[X^R]$. Thus it proves that the unified aggregate estimator of EJT developed in this paper works.

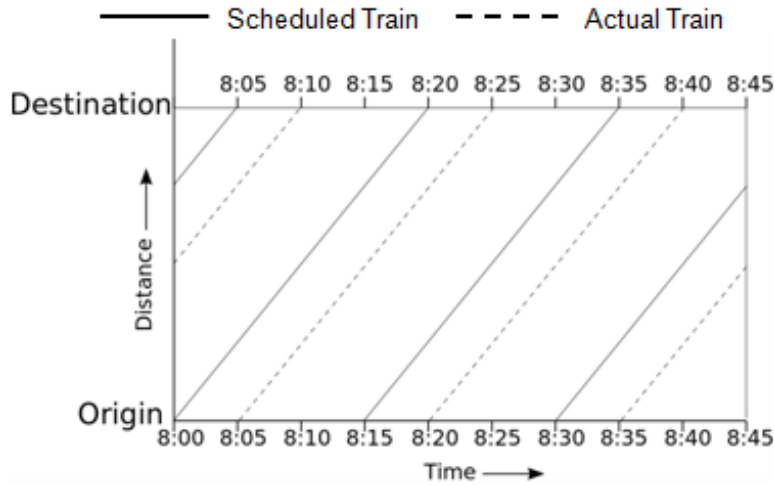


Figure 2. A perfect-headway example

4. Application to the London Overground network

This section presents EJT results for the London Overground network. It adopts the conceptual approach proposed in this paper and implements it using published Overground timetables and journey data from the Oyster smartcard ticketing system. It uses some of the results observed on the Overground as consideration points for discussing the properties and merits of EJT as a measure of service quality.

4.1. The London Overground and the oyster smartcard ticketing system

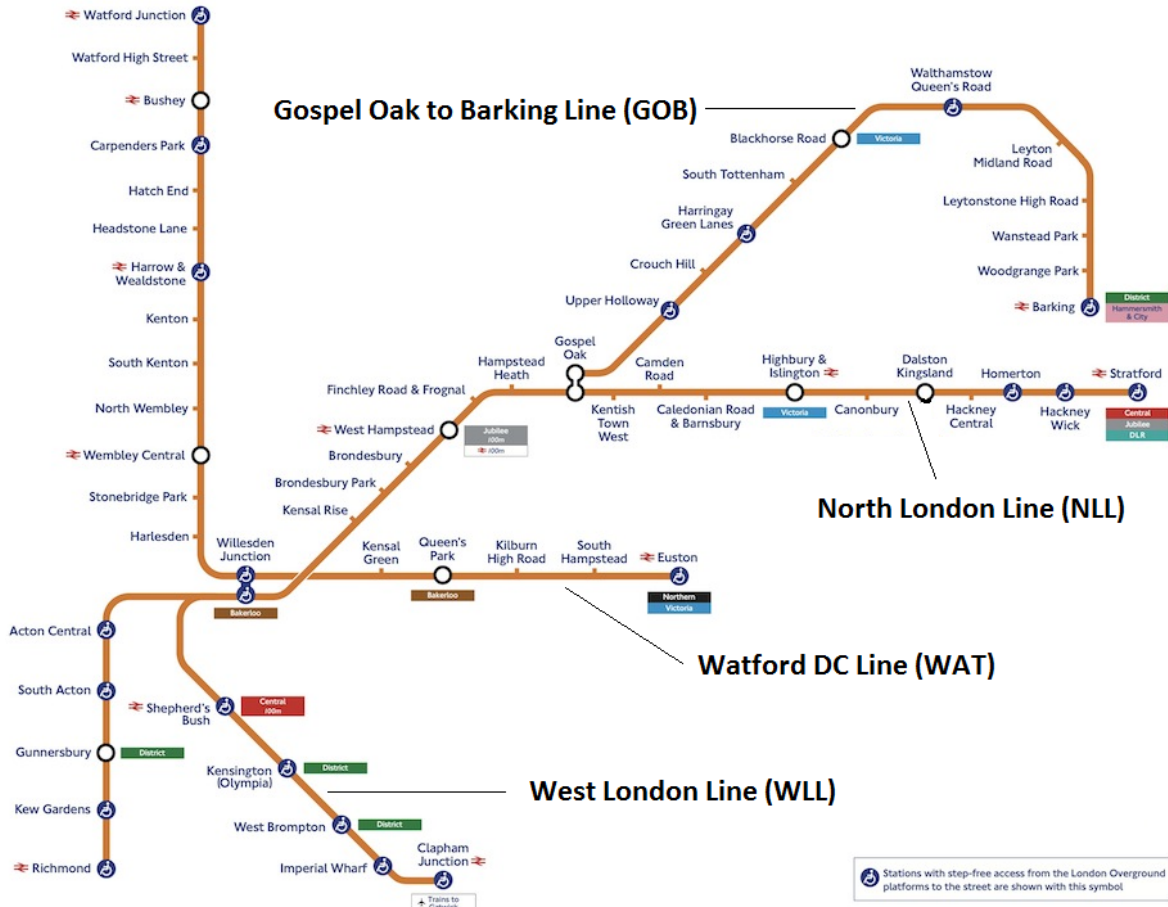


Figure 3. London Overground map (as of 2008)

The London Overground network is for the most part circumferential, primarily orbiting London to the North and West, and is very much part of the integrated network of Transport for London (TfL) and National Rail services. Services on the Overground are for the most part divided into four different service patterns: North London Line (NLL), Gospel Oak to Barking Line (GOB), Watford DC Line (WAT), and West London Line (WLL). NLL is the core of the network and the busiest Overground line, with the most frequent service. A map of the Overground network is shown in Figure 3.

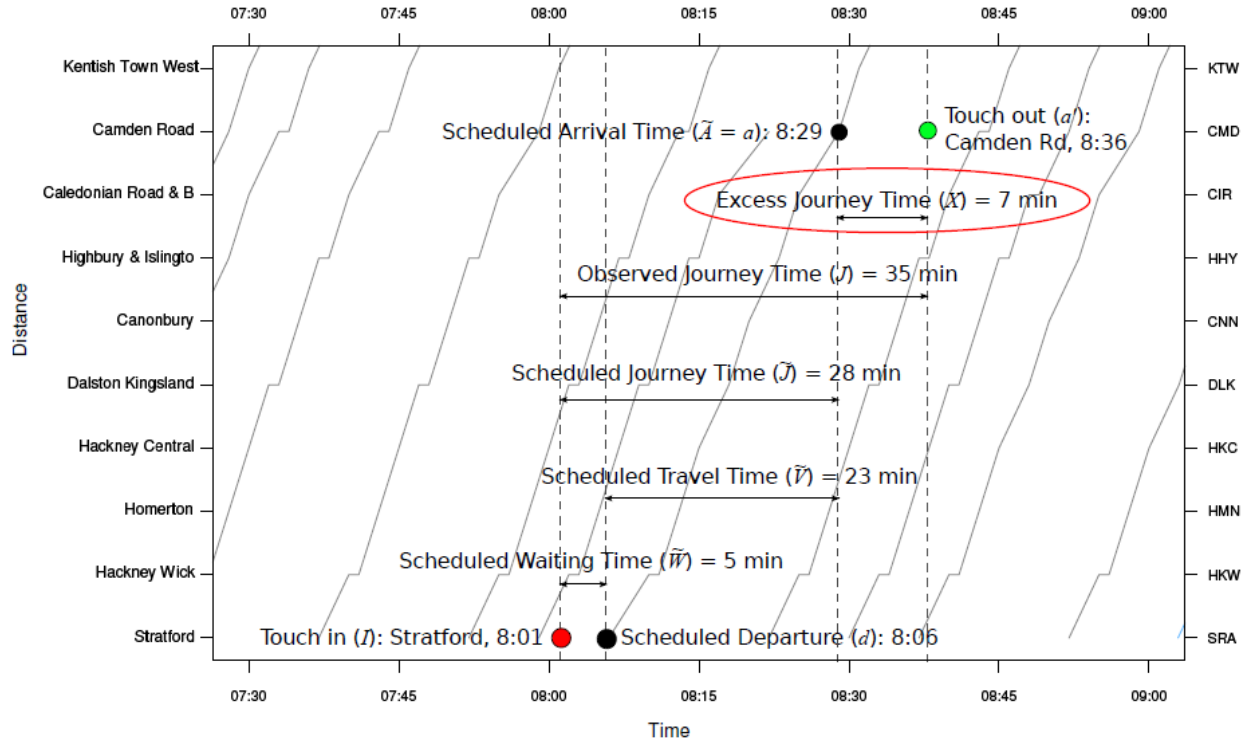
Oyster is TfL’s AFC smartcard system. London’s fare policy and technologies requires most Oyster users to validate their cards upon all entries and exits to the system. The centralized computer system archives these Oyster entry and exit transactions including their location, time stamp and Oyster ID in an easily accessible database. As a result, disaggregate Oyster journey data are cheap to gather in large volumes, and provide a prime source of data on individual passenger journey. Not everyone uses Oyster. The penetration rate across all TfL services is estimated to be approximately 80% (Transport for London, 2009), but varies in space and in time across the TfL network (Chan, 2007)

4.2. Implementation of the EJT methodology

EJT for London Overground journeys is estimated according to the unified methodology described in Section 3. This method assumes that the incidence behavior and journey time standards of all passengers are dependent on the timetable. This was shown to be unbiased in aggregate, even if some or all passengers are in fact randomly incident.

Under this framework, for each given journey recorded by the Oyster smartcard ticketing system

- 1 • the incidence time, I , is estimated as the timestamp of the entry transaction;
- 2 • the actual arrival time, a' , is estimated as the timestamp of the exit transaction;
- 3 • the scheduled arrival time, a , is estimated from the timetable;
- 4 • the total journey time, J , is estimated as $a' - I$ (the difference between the entry and exit transaction
- 5 times);
- 6 • the excess journey time, X , is estimated as $a' - a$ (the difference between the actual and scheduled
- 7 arrival times).
- 8

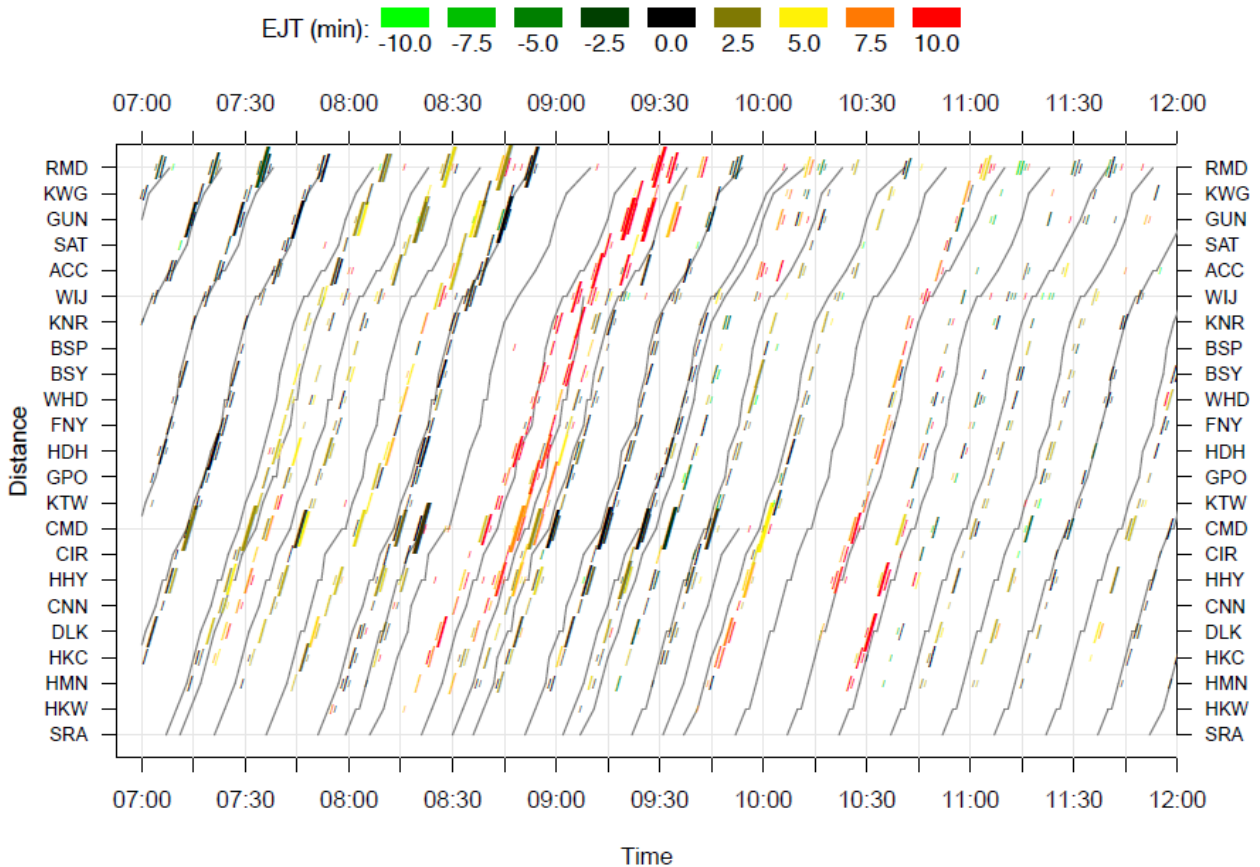


9
10 **Figure 4.** Time-distance illustration of EJT estimation for a passenger from Stratford to Camden Road.

11
12 Figure 4 illustrates this method for a passenger traveling from Stratford to Camden Road on the North
13 London Line (NLL) of the London Overground. In this “time-distance” plot, the X axis represents time
14 and the Y axis represents the distance traveled along the NLL in number of stations. Each line traveling
15 northeast through the plot shows the schedule of one weekday service.

16 The estimation of the scheduled arrival time, a , for each passenger is achieved through the same
17 schedule-based assignment process used to analyze passenger incidence behavior by Frumin and Zhao
18 (2012). In the algorithm a *Path()* function is used to encapsulate the complexity of conducting a schedule-
19 based assignment for a single passenger trip. This was implemented in the free/open source software
20 library Graphserver (Graphserver, 2009). Graphserver reads timetables in the widely used General Transit
21 Feed Specification (GIFS) (Google, 2009). This specification was defined by Google to facilitate transfer
22 of public transport schedules from operators to Google to power its own web-based journey planning
23 software. It has become a de-facto standard for public distribution of public transport timetables. A
24 combination of open source tools and scripts were used to process over 1.6 million Oyster journey records
25 made by over 290 thousand passengers on the 52 weekdays from 31 March, 2008 through 10 June, 2008,
26 inclusive. The data set was filtered to include only those journeys for which it is almost certain that the
27 passenger in question used only Overground services. The resulting data set contains nearly 1,670,000
28 journeys from 54 stations on 1,442 origin-destination pairs made by over 290,000 passengers. In that
29 sense, EJT is estimated through a two-stage assignment process. First, a frequency-based assignment is

1 used to select journeys that are (almost) certain to have used the Overground. Second, a schedule-based
 2 assignment is used to determine EJT with respect to the Overground timetable for those journeys.
 3



4
 5 **Figure 5.** Time-distance plot of timetable and observed Oyster exits for westbound travel on the North
 6 London Line on 3 April, 2008.
 7

8 Because of the complex and dynamic nature of even a single day's rail operations and passenger
 9 journeys, a graphical approach is used to validate the EJT measurements. Figure 5 shows the graphical
 10 validation results, and it can also help agencies to monitor service performance information. The plot is
 11 similar to that shown in Figure 4 with the addition of the times, locations, and EJT of actual passenger
 12 journeys. The size of the slanted hatch marks represents the number of Oyster journeys (in the Stratford –
 13 Richmond direction) that exited a given station on a given minute of the day on Thursday 3 April, 2008.
 14 The color of each hatch mark indicates the average EJT for the trips it represents. The more yellow and
 15 then red the mark, the more positive (i.e. late) the EJT; the greener the mark, the more negative (i.e. early)
 16 the EJT. There is much that can be inferred from this plot about the service delivery and quality on the
 17 day in question, for example:

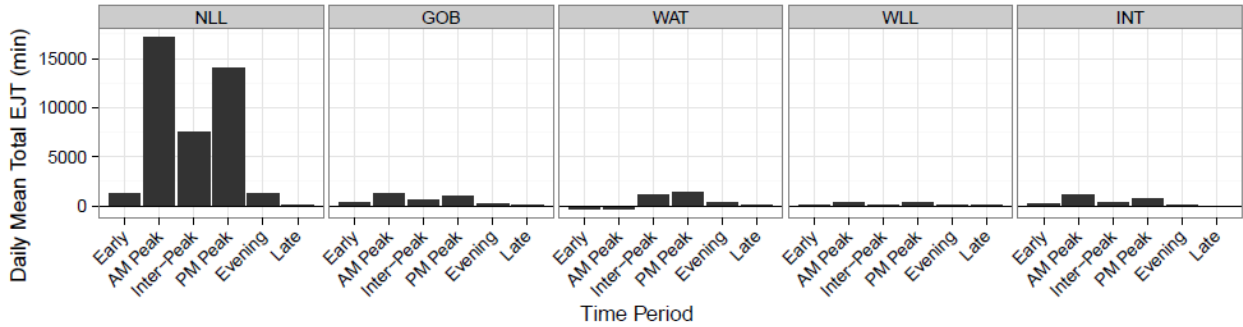
- 18 • Trains arriving Richmond (RMD) after 08:00 were less patronized and ran to schedule or a bit early.
- 19 • Starting with the 07:07 service from Stratford there are slight delays, which become severe for the
 20 08:06 and 08:22, and, perhaps, also the 08:30 and 08:37 trips.
- 21 • The 08:52, 09:03, 09:22, and 09:38 services ran smoothly, at least as far as Willesden Junction (WIJ).
- 22 • The 09:31 shuttle to Camden Road (CMD) may not have run at all, as reflected by the late passengers
 23 as far as Camden Road on the 09:38 service.
- 24 • The 09:52 from Stratford ran extremely late or not at all.
- 25 • By the 11:07 departure from Stratford, the service had largely recovered.

1 The above hypotheses were justified by the true record of events found in the Overground’s incident
 2 logs, and thus prove that that aggregate Oyster-based EJT measurements accurately reflect events on the
 3 ground, including train operations (service delivery) and the passenger's experience (service quality).

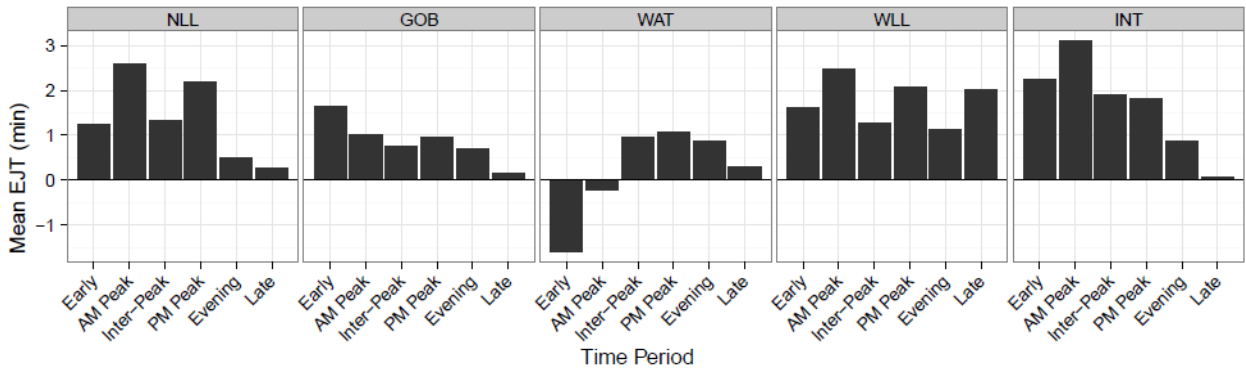
4
 5 **4.3. Results**

6
 7 This section presents EJT results for the London Overground network, first in isolation and next in
 8 comparison to the existing measure of service delivery.

9
 10 **4.3.1. Mean and total EJT by line and time period**



12
 13 **Figure 6. Total EJT, by line and time period**



15
 16 **Figure 7. Mean EJT, by line and time period**

17
 18 Figures 6 and 7 show two different aggregations of EJT by line and time period. The first of these
 19 presents the daily average of the sum of EJT for all Oyster journeys. This plot emphasizes the passenger-
 20 weighted nature of EJT as a measure of system performance. It is clear that, as a product of the number of
 21 passengers and the length of delays experienced by those passengers, the NLL, particularly during the
 22 AM and PM Peak periods, is the most problematic part of the Overground network. It is the line most
 23 deserving of management and tactical planning attention; the other lines frankly pale in comparison. For
 24 the Gospel Oak to Barking (GOB) line, the West London Line (WLL) and for interchange passengers
 25 (INT) the AM and PM Peak periods have the most total passenger delay. The Watford DC Line (WAT)
 26 breaks this pattern, with negative total and mean EJT in the Early and AM Peak periods.

27
 28 Figure 7 shows the pure mean passenger EJT. It puts all lines and time periods on equal footing by
 29 normalizing by the total number of journeys. This plot is primarily useful for comparing the performance
 30 of different lines at different times of day from the perspective of the average passenger, rather than all
 31 passengers. Overall mean EJT clearly varies across lines and time periods. After normalizing for the total
 32 number of journeys, the AM and PM Peak periods, with EJT of 2.6 and 2.2 minutes, respectively, are still
 the most problematic periods for the NLL.

1 These measurements are higher than all other lines for corresponding time periods except for
2 interchange passengers (INT) in the AM Peak, with an EJT of 3.1 minutes. It is not unexpected that
3 interchange passengers (most of whom likely use the NLL for one leg of their journey) suffer longer
4 delays than single-line passengers. A short delay on the first leg of an interchange journey can cause the
5 passenger to miss the targeted departure of the second leg. This could magnify the small delay on the first
6 leg to an entire headway of the service on the second leg.

7 The WLL is very close to the NLL in terms of mean EJT, whereas it was dwarfed in terms of total EJT.
8 The implication is that journeys on the WLL are subject to delays of similar (average) magnitudes, but
9 many fewer journeys are affected. Unlike on the NLL, the normalization by total passengers does change
10 the relative picture for the GOB. While total EJT is greatest in the AM and PM Peak periods, the highest
11 average EJT is experienced by passengers during the Early Morning period. These relative differences do
12 not necessarily mean that one time period or line is more worthy of attention than the other. Rather, it
13 presents a more nuanced picture of service quality which can be acted upon differently depending on
14 management policies and priorities.

15 EJT is net negative on the WAT in the Early and AM Peak periods. To understand this further, EJT of
16 WAT passengers were investigated at the level of individual origin-destination (OD) flows. In the AM
17 Peak, 46 out of 236 OD flows on the WAT had net negative EJT, including all 15 OD flows into Euston,
18 the line's southern central London terminus. These flows into Euston account for 93% of net negative EJT
19 on the 46 net negative OD flows. Almost half of that net negative EJT into Euston comes from the flows
20 originating at Queens Park and Kensal Green (towards the southern end of the line), both with average
21 EJT of almost -3.0 minutes. This is explained, in consultation with Overground management (Bratton,
22 2008) by the fact that WAT trains often depart Queens Park on time or slightly late and arrive Euston
23 terminal up to 5 minutes early. In other words, their scheduled running time over the last portion of the
24 line is generous.

25 Another quarter of the net negative EJT into Euston comes from the OD flows originating between
26 Watford Junction and Harrow & Wealdstone (at the northern end of the line), non-inclusive, which have a
27 mean EJT of -5.1 to -12.4 minutes. This likely represents a problem with the assignment model used to
28 filter non-Overground passengers. Another National Rail service provides twice-hourly express service
29 from Watford Junction and Harrow & Wealdstone to Euston in substantially less time than the
30 Overground. The assignment model correctly assigns passengers from these two stations to that service,
31 but not for passengers who start on the Overground and interchange to this express service, perhaps
32 opportunistically, at Harrow & Wealdstone.

33 In general, these aggregate results are in line with a priori expectations held by the management of the
34 Overground network (e.g. Bratton, 2008). The most strongly held belief, confirmed here, is that the NLL
35 carries the largest passenger loads and has the most delays, especially during the peak periods.

36 37 *4.3.2. Time series of mean EJT*

38 Figure 8 shows daily mean EJT over time for each London Overground line, for the whole day and for
39 the AM Peak period. On all lines, there is marked day-to-day and week-to-week variability of EJT. As
40 expected, mean EJT exhibits some volatility on a day-to-day basis. This is particularly the case as sample
41 sizes decrease, namely in the AM Peak compared to the whole day, and for the WLL and INT compared
42 to the other lines. There is not a clear up or down trend over time in this dataset suggesting that overall
43 relative service quality on the Overground was steady over this period.

44

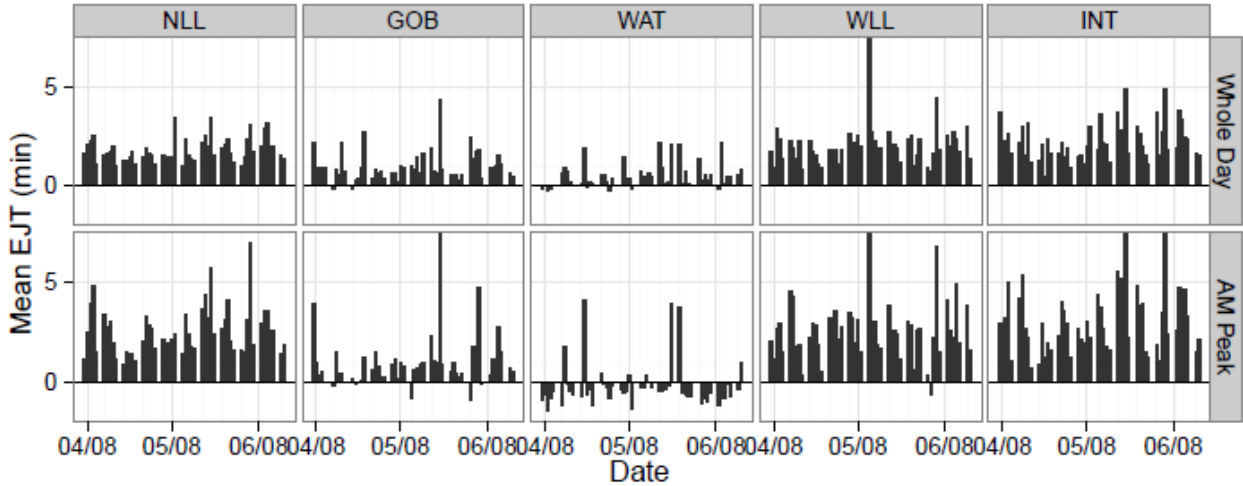


Figure 8. Daily mean EJT.

4.3.3. Mean and total EJT by time period and direction (NLL)

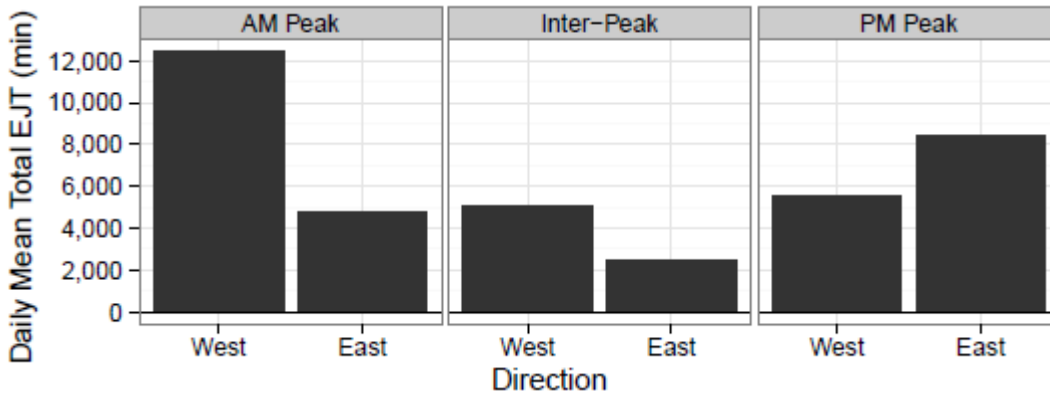


Figure 9. Total EJT on the NLL, by time period and direction.

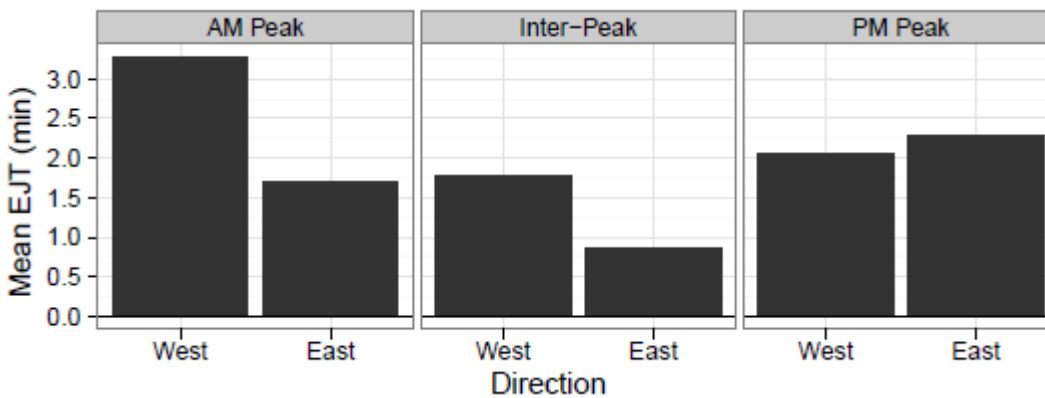


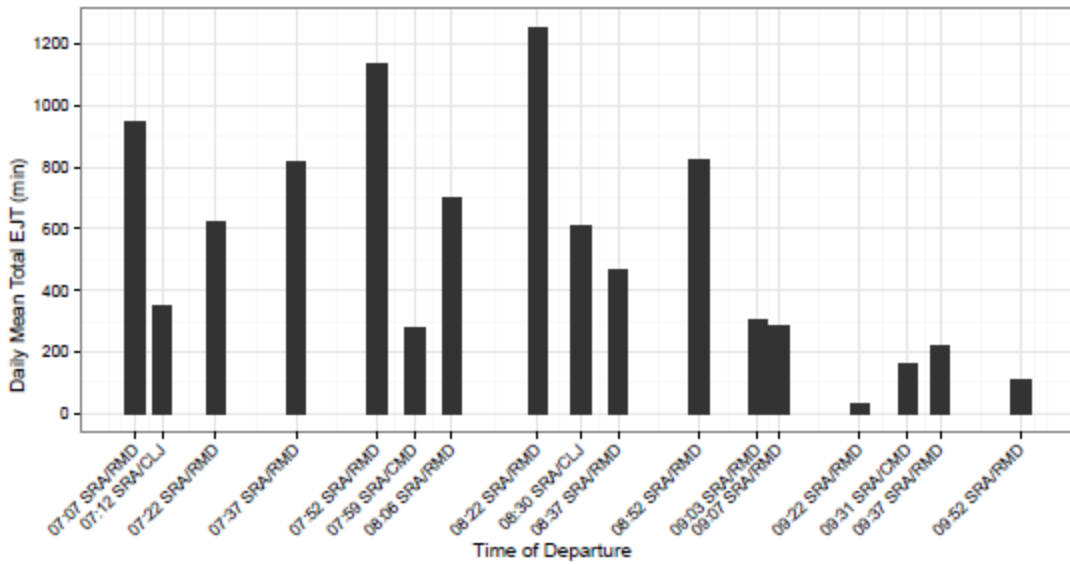
Figure 10. Mean EJT on the NLL, by time period and direction.

Figures 9 and 10 further disaggregate EJT results for the NLL by direction of travel. This aggregation is important because of the unbalanced nature of passenger demand on the NLL (and indeed on many railways) in different periods of the day. Figure 9 shows total EJT to be substantially worse in the westbound direction than in the eastbound direction in the AM Peak period. Figure 10 shows mean EJT to

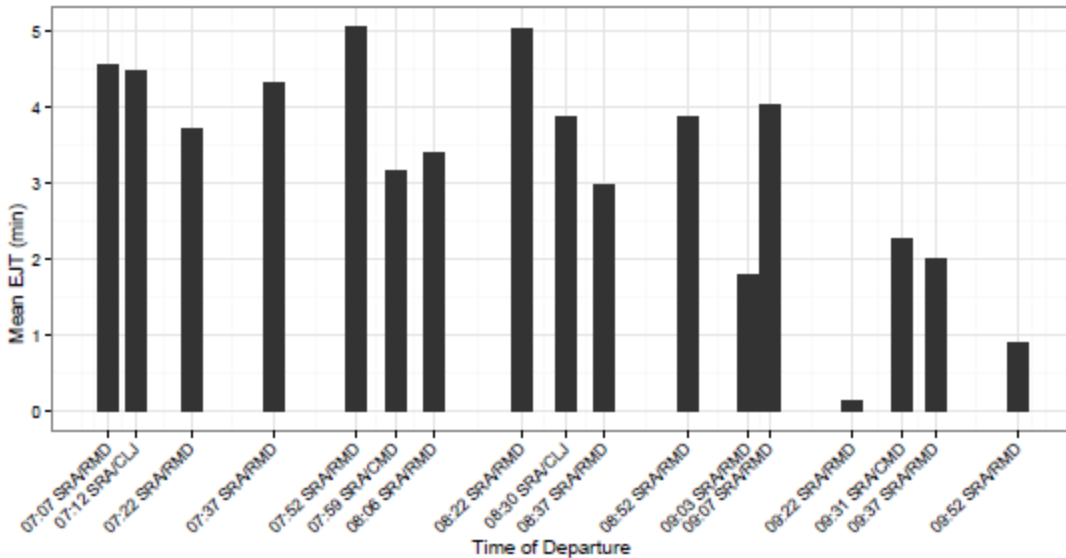
1 be similarly unbalanced, though somewhat less so than total EJT, in the same period. This indicates that,
 2 in the AM Peak period, there are more passengers suffering longer delays in the westbound than the
 3 eastbound direction. Similar results can be seen in the PM Peak period with the directions reversed,
 4 though the unbalance is not nearly as severe as in the AM Peak.

5
 6 *4.3.4. Mean and total EJT by scheduled service (NLL AM peak)*

7 One advantage of the approach proposed here is that, with such a large and detailed data set, it is
 8 possible to probe deeper into the specifics of delays and their effects on passengers. To estimate EJT,
 9 each passenger journey was assigned to a specific scheduled service (Frumin and Zhao, 2012). This
 10 assignment indicates only which train a given journey would likely have taken under right-time service
 11 delivery, not which train the passenger actually rode. In that sense, each scheduled service defines a
 12 specific market over time and space, and the assignment places journeys into these markets.
 13



14 **Figure 11.** Total EJT by scheduled service, westbound.



17 **Figure 12.** Mean EJT by scheduled service, westbound.

1 Figures 11 and 12 aggregate EJT to the level of these markets. They show total and mean EJT,
 2 respectively, for westbound passenger journeys on the NLL between Stratford and Willesden Junction.
 3 These are the peak London Overground markets - the peak direction (from Stratford towards Richmond)
 4 at the peak time of day on the busiest line - which were shown in Figures 5 and 6 to have the most severe
 5 EJT problems on the whole Overground network. The bars in these plots are spaced according to the
 6 actual headway (at Stratford) of each service.

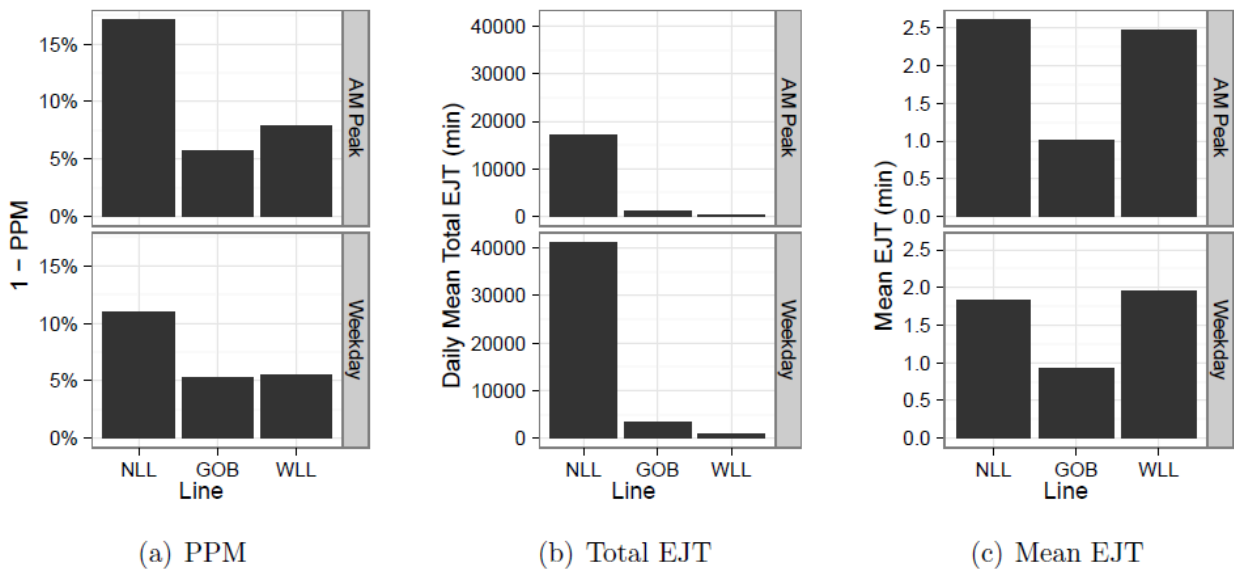
7 Figure 11 clearly shows that the 07:52 and 08:22 trains are the most problematic services in terms of
 8 total passenger delay. It also shows how unbalanced the headways in the timetable are, especially
 9 between Stratford (SRA) and Camden Road (CMD). Between 07:00 and 09:00, the services to Richmond
 10 with full 15 minute headways have the highest total EJT relative to their shorter-headway leaders and/or
 11 followers.

12 The mean EJT results in Figure 12, normalized by the total number of journeys in each market, show
 13 similar results. The most substantial relative differences are for the short headway services at 07:12 and
 14 09:07. Their mean EJT is much higher compared to other services than was their total EJT. This stands to
 15 reason, in that with short headways they should have fewer journeys and thus less total EJT.

16
 17 *4.3.5. Comparison with existing performance metrics*

18 EJT measures are compared with corresponding on-time performance (OTP) results from the existing
 19 London Overground performance regime, the Public Performance Measure (PPM). These comparisons
 20 are presented as much to explore the differences between EJT and OTP as public transport performance
 21 measures as to characterize performance on the Overground.

22 Figure 13 plots PPM and total and mean EJT by line, for the AM Peak and for the whole day, over the
 23 entire study period. The plot shows the complement of PPM so that the measures are directionally aligned
 24 (i.e. a higher number indicates worse performance). PPM and total EJT correspond in that the NLL is by
 25 far the worst performing line. The difference between the NLL and other lines is even more pronounced
 26 in terms of total EJT than in terms of PPM. This reflects the difference in passenger volumes between the
 27 different Overground lines.
 28



29
 30 **Figure 13.** EJT and PPM, by line.
 31

32 The WLL appears much worse than the other lines in terms of mean EJT than it is in terms of PPM.
 33 This could be because the WLL has the lowest frequency of the Overground lines. Consequently, each
 34 delayed train (counting against PPM) may have a greater proportionate effect on the line's passengers. It

1 could also be that the manner in which WLL trains are delayed has worse effects on passengers than on
2 the other lines.

3 4 5 **5. Discussion and conclusion**

6
7 This section anticipates and discusses some concerns that may arise in the application of the method
8 proposed in this paper. And in the end, a few conclusions are drawn.

9 10 *5.1. Application considerations*

11
12 AFC penetration rates may vary across the network for which EJT is measured. In some cases, this
13 may require a weighting of EJT values to account for this variation. For example, if the rate varies
14 significantly across different OD flows on the same line, re-weighting may be needed when analyzing
15 EJT on that line. If penetration is largely consistent for a given line but varies across line, such correction
16 is only necessary if comparing EJT results between those lines.

17
18 Care must be taken if EJT is to be measured for only some portion (e.g. the Overground) of a broader
19 network (e.g. the entire London railway network), especially if timetables are available only for that
20 portion of the network. The most straightforward way to handle this situation is to select only those OD
21 flows that will use the portion of the network in question with relative certainty. This can be done
22 manually based on judgment, or can take advantage of an assignment model that considers the entire
23 network.

24 25 *5.2. Negative EJT*

26
27 EJT for an individual passenger journey (under the scheduled incidence assumptions) can be negative.
28 This is in and of itself not a cause for concern in terms of the measurement of EJT. However, aggregate
29 EJT that is net negative may indicate certain biases in the EJT estimation process. Negative EJT for
30 individual journeys can occur for several reasons, including the following.

- 31 • A passenger uses some service not included in the set of timetables used in setting journey time
32 standards. In this case, the negative EJT can be smaller or larger in magnitude than the headway of
33 the service in question.
- 34 • A passenger takes the service on which he or she is scheduled to depart, but that service arrives at that
35 passenger's destination earlier than the timetable indicates. In this case, when the headway is
36 relatively large, the negative EJT should be small in magnitude relative to the headway of the service
37 in question.
- 38 • A passenger takes an earlier service than the one which he or she is scheduled to depart (because that
39 earlier service was running late), which naturally arrives at the passenger's destination before the
40 passenger's scheduled arrival time. In this case, the negative EJT can be as large as the headway of
41 the service in question.

42 The first of these reasons indicates a potential bias the estimation of EJT. If non-scheduled trips were
43 inserted into the timetable by the operator in question as a result of service control decisions, the negative
44 EJT is unbiased in that passengers would not be considered to have expected to use this new service.
45 However, if the service that the passenger used was provided by a different operator (e.g. one who shares
46 the same track, or on a different path altogether), the result can be considered a biased EJT in that the
47 service should have been used in setting journey time standards. This reflects a problem with the selection
48 of OD flows for which EJT is measured for a given operator, which may result from biases in an
49 assignment model used to select those OD flows. The second and third of these reasons do affect the EJT
50 estimate for an individual journey. These negative EJT measurements clearly affect the distribution of
51 EJT, but should not under most circumstances unduly affect the mean.

5.3. EJT and longitudinal analysis

A problem for using EJT in longitudinal analyses is that when the timetable is revised, changes in EJT may be driven more by the timetable modification than by any real changes in journey times experienced by passengers. For example, if running times in the timetable are lengthened but passenger journey times remain the same, EJT will decrease even if passengers experience no actual improvement in actual journey time.

Furthermore, passengers may adjust their incidence behaviour over time as service conditions change. The method proposed here is entirely conditioned on actual incidence behavior, so changes in such behavior will not bias estimates of EJT per se. In one sense, this is a positive feature of this method because it absolves the analyst of the need to make any assumptions regarding incidence behavior. However, it also implies that EJT will not capture some of the benefits of improved service reliability. Specifically, it will not reflect the benefits captured by passengers who take advantage of more reliable service by adjusting their incidence behavior to reduce waiting time (likewise for the harm to passengers who react to less reliable service by becoming more randomly incident incidence).

For example, consider a service that has become more reliable over time, perhaps because of improved infrastructure or rolling stock but with no changes to the timetable. If, as a result of this reliability improvement, passengers of this service now arrive at their destinations closer to their respective arrival time standards, such will be reflected in EJT measurements. However, it may also be the case that the journey time standards of some of these passengers has decreased because, as the service has become more reliable, their incidence behavior has become less random (i.e. more timetable-dependent, with smaller scheduled waiting times). This would not be reflected in EJT measurements.

These realizations highlight the relative nature of EJT, and suggest that other measures, for example those for absolute service quality, may be necessary for longitudinal analysis. It should also be noted that measures of service quality, including EJT, are not intended to be used for evaluating a timetable on its own merits. They simply speak to the differences between passengers' actual journeys and the promise implied by the timetable. Evaluation of a timetable in isolation from passenger journeys is not considered here.

5.4. Extension to a heterogeneous rail network with interchanges

The authors' intuition is that this analysis extends readily beyond a single rail line with a single service pattern to a rail network with interchanges and a variety of service patterns. Such an extension would require the model to account for passenger incidence behavior at interchange locations. Without formally extending the model, the following observations should provide an intuitive sense of why the schedule-based estimator for aggregate EJT is appropriate in a network context. Note that in all cases it is expected to measure only the end-to-end journey time, which subsumes all interchanges.

- This analysis easily extends to include a network with walking links, such as those between AFC gatelines and station platforms. Such links can be thought of as lines or services with continuously available departure times (i.e. infinite frequency, or zero headway), in which case the distinction between scheduled and random incidence is irrelevant.
- On a single line with heterogeneous service patterns, such as short turns or a trunk-and-branch configuration, passengers can be considered to ignore certain departures that do not improve their overall travel time (Frumin and Zhao, 2012). This simply changes the timetable applicable to each passenger's journey, not the analysis thereof.
- If passengers are aware of and make plans based on the timetable for the entire network, then clearly the schedule-based estimator is appropriate.
- If passengers are unaware of or do not use the timetable for any portion of the network, and if the different services are timetabled independently, then incidence at the interchange location will be

1 random with respect to the departures of the service being interchanged to. In this case, some
 2 passengers will be lucky and experience short interchange times while others will experience long
 3 ones.

- 4 • If the timetable is designed to facilitate interchanges (i.e. minimize interchange times) between lines,
 5 then a schedule-based estimator should be used regardless of passenger incidence behavior and
 6 standards. Even if passengers are randomly incident at the initial station, their incidence at the
 7 interchange station is non-random with respect to the timetable of the subsequent line by the very
 8 nature of the specially-constructed timetable.
- 9 • If passengers are aware of and use the timetable for only a portion of the network, then they either
 10 interchange from a service on which they schedule their incidence to a service on which they are
 11 randomly incident, or vice versa. In the former case, since they do not know (or care) about the
 12 timetable for the second service, their incidence time (and thus journey time standard) for the first
 13 service is unaffected.
- 14 • The reverse scenario, where passengers are unaware of the timetable on their first service, but have a
 15 target departure in mind for the second, is perhaps less straight-forward. In this case, it would be
 16 reasonable to set a waiting time standard of a full (rather than half) headway for the first service, since
 17 this is what an operator would recommend, based on the timetable, to minimize the probability of
 18 missing the second, scheduled, departure. A first intuition is that this would bias some of the analyses
 19 in this paper. However, as was found in those analyses, the first intuition with respect to incidence
 20 behavior, the timetable, and journey time standards is not always correct. This issue merits further
 21 examination.
- 22 • If these intuitions are to be believed, and the issue in the final observation is resolved, the model and
 23 estimators for aggregate EJT developed in this paper are in fact quite general and should be
 24 applicable to a wide variety of contexts.

25 26 *5.5. Conclusions*

27
28 Excess journey time (EJT), with standards derived from the timetable, is a measure of relative service
 29 quality that strikes useful balance between the passenger's and operator's perspectives. It has found lasting
 30 application at a number of large urban railways. Actual passenger journey times can now be measured
 31 (rather than modeled) directly from automatic data produced by AFC systems such as the Oyster
 32 smartcard.

33 Along with measuring actual passenger journey times, EJT depends on a standard against which to
 34 compare those measurements. These standards should be based on the timetable, so as to be as
 35 meaningful and useful to operators as possible. Within that constraint, they should reflect passenger
 36 concerns as realistically as possible. Most measures of service quality and relative service quality have
 37 made the assumption of random incidence, with the implied standard for waiting time of half the
 38 scheduled headway. Passenger incidence behavior is often, including on the London Overground, much
 39 more heterogeneous than that. This heterogeneity of the behavior comes with certain implications about
 40 what knowledge passengers have of the timetable and how they use that knowledge.

41 Based on this, this paper has established a rigorous framework for analyzing EJT, in particular for
 42 reasoning about passenger' journey time standards as implied by varying incidence behaviors. It was
 43 found that the wrong assumption about incidence behavior and journey time standards can result in a
 44 biased estimate of EJT at the level of an individual passenger journey. Nevertheless, the estimator for
 45 aggregate EJT is unbiased and unified, regardless of actual passenger behavior, under the assumption that
 46 all passenger incidence and associated journey time standards are dependent on actual departure times in
 47 the timetable. This result was proven for a single rail line without interchanges, but intuitively should
 48 hold for a rail network. This is a very useful result in practice. It allows for the estimation of aggregate
 49 EJT from only AFC (e.g. Oyster smartcard) data and published timetables in a simple unified manner,
 50 regardless of service frequencies or passenger behaviors that vary across the network or over time.

1 The paper also presents an analysis of aggregate and disaggregate EJT results for the London
 2 Overground, both in isolation and in comparison to the Overground's existing measure of service delivery.
 3 EJT for individual passenger journeys on a given service was found to range from negative (i.e. early) by
 4 up to one headway to positive (i.e. late) by substantially more than one headway. However, it is difficult
 5 to interpret EJT for individual journeys, in part because of the ambiguity with respect to passenger's
 6 standards and incidence behavior discussed. Consequently, EJT is not a particularly useful measure to
 7 analyze individual passenger journeys. Aggregate EJT, on the other hand, is a measure of relative service
 8 quality with clear meaning. It expresses the average passenger's experience in terms of total journey time
 9 compared to what the timetable would imply, for a wide range of passenger incidence behaviors.
 10 Individual EJT measurements are, by nature of the assignment process by which they are estimated, easily
 11 aggregated both spatially and temporally. Depending on the analytical need, aggregate EJT can be
 12 estimated at the level of line, origin-destination flow, scheduled service, time period (e.g. AM Peak), day,
 13 week, etc. Aggregate EJT was found to vary substantially across the different London Overground lines
 14 and across time periods of weekday service. Total EJT is greatest on the North London Line in the AM
 15 and PM Peak periods, which also had among the highest estimates of mean EJT.

18 Acknowledgement

19
 20 This paper is partly based on the second author's master thesis in Massachusetts Institute of
 21 Technology (Frumin, 2010). The authors thank Transport for London for providing the financial and
 22 institutional support for this research, the opportunity to work in the agency and the access to the data
 23 used in this paper, and specifically Oliver Bratton at London Overground Rail Operations LTD for insight
 24 into the issues and dynamics of the London Overground network.

27 References

- 28 Abkowitz, M., Slavin, H., English, L. and Wilson, Nigel H. M. (1978) *Transit Service Reliability*, Cambridge, MA.
 29 Bates, John, Polak, John, Jones, Peter and Cook, Andrew (2001) 'The valuation of reliability for personal travel',
 30 *Transportation Research Part E: Logistics and Transportation Review*, 37, pp. 191–229, [online] Available
 31 from: <http://linkinghub.elsevier.com/retrieve/pii/S1366554500000119>.
 32 Bratton, Oliver (2008) 'Unpublished interviews and electronic communications with Oliver Bratton, Deputy
 33 Operations Director, London Overground Rail Operations LTD'.
 34 Buneman, K. (1984) 'Automated and passenger-based transit performance measures', *Transportation Research*
 35 *Record*, (992), pp. 23–28.
 36 Camilletti, P. (1998) *High Frequency Calculations Explained, London Buses*,.
 37 Camilletti, P. (2003) *Low Frequency Calculations Explained, London Buses*,.
 38 Chan, Joanne (2007) *Rail Transit OD Matrix Estimation and Journey Time Reliability Metrics Using Automated*
 39 *Fare Data, Matrix*, Master thesis, Massachusetts Institute of Technology.
 40 Friedman, R. R. (1976) 'Statistical models of the mean and standard deviation of passenger wait time in urban bus
 41 transit', Northwestern University.
 42 Frumin, Michael (2010) *Automatic Data for Applied Railway Management: Passenger Demand, Service Quality*
 43 *Measurement, and Tactical Planning on the London Overground Network*, Master Thesis, Master thesis,
 44 Massachusetts Institute of Technology.
 45 Frumin, Michael and Zhao, Jinhua (2012) 'Analyzing passenger incidence behavior in heterogeneous transit services
 46 using smartcard data and schedule-based assignment', *Transportation Research Record: Journal of the*
 47 *Transportation Research Board (Accepted)*.
 48 Furth, Peter, Hemily, T. and Strathman, J. (2006) *Using Archived AVL-APC Data to Improve Transit Performance*
 49 *and Management*, Transportation Research Board, Washington D.C.
 50 Furth, Peter and Muller, Theo (2006) 'Part 4: Capacity and Quality of Service: Service Reliability and Hidden
 51 Waiting Time: Insights from Automatic Vehicle Location Data', *Transportation Research Record*, 1955(1),
 52 pp. 79–87, [online] Available from:
 53 <http://trb.metapress.com/openurl.asp?genre=article&id=doi:10.3141/1955-10>.

- 1 Furth, Peter and Muller, Theo (2007) 'Service reliability and optimal running time schedules', *Transportation*
2 *Research Record*., pp. 55–61.
- 3 Google (2009) 'General Transit Feed Specification', [online] Available from:
4 http://code.google.com/transit/spec/transit_feed_specification.html.
- 5 Graphserver (2009) 'Graphserver: The Open-Source Multi-Modal Trip Planner', [online] Available from:
6 <http://bmander.github.com/graphserver/>.
- 7 Hammerle, M., Haynes, M. and McNeil, S. (2005) 'Use of automatic vehicle location and passenger count data to
8 evaluate bus operations: experience of the Chicago transit authority', *Transportation Research Record*, 1903,
9 pp. 27–34.
- 10 Henderson, G., Adkins, H. and Kwong, P. (1991) 'Subway reliability and the odds of getting there on time',
11 *Transportation Research Record*, (1297), pp. 10–13.
- 12 Henderson, G., Adkins, H. and Kwong, P. (1990) 'Toward a passenger-oriented model of subway performance',
13 *Transportation Research Record*, (1266), pp. 221–228.
- 14 Henderson, G., Kwong, P. and Heba, A. (1991) 'Regularity indices for evaluating transit performance',
15 *Transportation Research Record*, 1297, pp. 3–9.
- 16 Kittelson & Associates (2003) *A Guidebook for Developing a Transit Performance-Measurement System*,
17 Transportation Research Board, Washington D.C.
- 18 Larson, R. C. and Odoni, A. R. (2007) *Urban Operations Research*, 2nd ed. Dynamic Ideas.
- 19 London Transport (1999) *Journey Time Metric (JTM): An Overview*,.
- 20 Osuna, E. E. and Newell, G. F. (1972) 'Control strategies for an idealized public transportation system',
21 *Transportation Science*, 6(1), pp. 52–71.
- 22 Reddy, A., Crammer, A., Cucarese, J., Lu, A. and Tran, M. (2009) 'Performance measurements on mass transit -
23 New York city transit authority case study', *Transportation Research Record*, 2111, pp. 125–138.
- 24 Transport for London (2009) *Oyster Factsheet*, [online] Available from:
25 <http://www.tfl.gov.uk/assets/downloads/corporate/oyster-factsheet-july-2009.pdf> (Accessed February 8, 2010).
- 26 Uniman, David Louis (2009) 'Service Reliability Measurement Framework using Smart Card Data: Application to
27 the London Underground', *Transport*, Massachusetts Institute of Technology.
- 28 Wilson, Nigel H. M., Attanucci, J. P., Chan, Joanne and Cui, A. (2008) 'The use of automated data collection
29 system to improve public transport performance', In *Proceedings of the 10th International Conference on*
30 *Application of Advanced Technologies in Transportation*.
- 31 Wilson, Nigel H. M., Nelson, D., Palmere, A., Grayson, T. and Cederquist, C. (1992) 'Service-Quality Monitoring
32 for High-Frequency Transit Lines', *Transportation Research Record*, pp. 3–11.
- 33