Service Reliability and Hidden Waiting Time:

Insights from AVL Data

Peter G. Furth (corresponding author)
Department of Civil & Environmental Engineering
Northeastern University
Boston, MA 02115
e-mail: pfurth@coe.neu.edu
telephone: 617.373.2447; fax: 617.373.4419

Theo H.J. Muller
Delft University of Technology
Faculty of Civil Engineering and Geosciences
Transportation and Planning Section
Postbus 5048
2600 GA Delft
The Netherlands
e-mail: Theo.Muller@citg.tudelft.nl
telephone: +31 15.278.5288; fax: +31 15.278.3179

Submitted for publication in *Transportation Research Record*
March 31, 2006

Word count: 5636 + 7 figures/tables @ 250 = 7386

ABSTRACT: Traditional transit service quality measures separate waiting time from service reliability, thereby underestimating the real cost of waiting and failing to evaluate the impact of unreliability on passengers. Analyzing passenger behavior, we show that for short headway service, the cost of waiting involves not only the mean time spent waiting on the platform, but also "potential waiting time," the additional time that passengers have to budget for waiting. Budgeted waiting time is based on an extreme of the waiting time distribution such as its 95-percentile value, which is extremely sensitive to service reliability. Methods for determining the distribution of passenger waiting time from automatic vehicle location (AVL) data are derived. For long headway service, actual and budgeted waiting time are shown to be related to high and low extremes of the schedule deviation distribution, which can likewise be determined from AVL data. Two other components of long headway waiting, schedule inconvenience and synchronization cost, are also analyzed. Waiting cost functions and waiting time measures that account for both headway and service reliability are developed, and are harmonized in a framework that provides a smooth transition from short- to long-headway waiting. Examples show how service reliability can be measured as a waiting cost, and how service reliability improvements can reduce waiting cost as much as a large reduction in headway.

Waiting time is an important deterrent to using transit, and therefore an important determinant of both demand and service quality. Predicting and monitoring passenger waiting time is therefore important for planning and management. Unfortunately, direct measures are not feasible because transit passengers are not individually tracked arriving and departing at stations. As part of TCRP project H-28 concerning the use of archived automatic vehicle location (AVL) data to improve transit management and performance (1), we have investigated how this detailed, though still indirect, source of data might offer new insights into passenger waiting.

The importance of good performance measures to strategic management is underscored by recent publications (2, 3). However, traditional measures of waiting time tend to focus on mean values, while passengers' perceptions tend to be based on extreme values, which are highly dependent on service reliability. At the same time, traditional measures of service reliability (*e.g.*, headway coefficient of variation (*cv*), on-time performance) are indicators of operational quality; they do not express reliability's impact on passengers. Consequently, waiting time is underestimated, and service reliability is undervalued.

Our objective is to develop measures and models that properly account for the impacts of service reliability on waiting-related user cost. For example, while everybody agrees that improving on-time performance from 80% to 90% is a good thing, we lack methods to estimate the resulting impact on passengers in standard units, which planners could then translate into an economic impact or a ridership forecast.

First short headway service is analyzed, then long. For each, we develop a waiting cost model, key measures of waiting time, and example reporting formats. The short- and long-headway waiting cost models are then harmonized. The paper ends by dealing with some practical issues and offering conclusions.

SHORT HEADWAY WAITING

When headways are short, passengers can be assumed to arrive independent of vehicle arrivals. This assumption, together with the assumption that passengers can always board the first vehicle (*i.e.*, no pass-ups), is sufficient to derive a distribution of passenger waiting time from the distribution of headways.

Mean Platform Time

Passengers' most obvious burden related to waiting is "platform time," the time spent waiting at the station. It is the time between a passenger's arrival and the next vehicle departure. For simplicity, we use the term "bus" to represent any transit vehicle.

Let H be the bus headway at a given station, *i.e.*, the time between its departure and the previous bus's departure. Passengers arriving during that headway have a uniform waiting time distribution on the interval $[0, H]$. If $H_{schedule}$ is the mean scheduled headway, and every departure in a given period has the same headway $h = H_{schedule}$, the waiting time distribution for the period as whole is uniform on the interval $[0, H_{schedule}]$. This is the basis of the nominal waiting time commonly used in planning applications,

$$W_{nominal} = 0.5\, H_{schedule} \tag{1}$$

Of course, headways are not all equal. Passengers arriving during long headways will, on average, have longer waits than those arriving during short headways. And because passengers are more likely to arrive during a long headway than during a short headway, mean waiting time is skewed toward that of the longer headways, and is consequently greater than 0.5 E[H]. One of the classic results of transportation science is the formula for mean waiting time (4):

$$\mathrm{E}[W] = 0.5\mathrm{E}[H]\left(1 + cv_H^2\right) \tag{2}$$

where W = waiting time and cv_H = headway cv.

Through this formula, a relationship between service reliability and average waiting time has been known for many years. However, as we will show later, average waiting time is not a measure that adequately reflects passengers' waiting cost, and therefore this formula falls short of accounting for the impact of reliability on waiting time.

Ideal and Excess Waiting Time

The mean waiting time formula is a product of two terms. The first (E[H]) is often seen as a result of planning, and the other a result of operations. In fact, however, planning and operations affect both terms, since planned headways sometimes have variations, and operations, through missed, added, and delayed trips, can result in a mean headway that differs (particularly in a prescribed period) from what was planned.

Still, the concept of separating the incremental impact of operations from that of planning is a useful framework. Following Wilson *et. al.* (5), mean waiting time can be partitioned into *ideal* and *excess mean waiting time*:

$$W_{ideal} = 0.5\,\mathrm{E}[H_{schedule}]\left(1 + cv_{H,schedule}^2\right)$$

$$W_{excess} = \mathrm{E}[W] - W_{ideal} \tag{3}$$

where $cv_{H,schedule}$ is the cv of scheduled headways. In London, where a public agency establishes headways while private contractors operate bus service, excess mean waiting time is used as an incentive in service contracts. Note that by adding extra trips, excess waiting time can be negative.

Distribution of Waiting Time

Under the same assumptions behind the mean waiting time formula (passengers arrive independent of the timetable, and no pass-ups), the distribution of waiting time can be derived from the headway distribution. Let

$f_H(h)$ = probability density of headways, distributed over bus trips

$f_{H,pax}(h)$ = probability density of headways, distributed over passengers

$f_W(w)$ = probability density of passenger waiting time, distributed over passengers

The first density relates to the probability that a randomly selected *trip* will have a headway h; the second, to the probability that a randomly selected *passenger* arrives during a headway h. Newell (4) showed that

$$f_{H,pax}(h) = \frac{h}{\mathrm{E}[H]} f_H(h) \tag{4}$$

In words, the probability that a passenger arrives during a headway of length h is proportional both to the popularity of that length, and to the length itself, since longer headways serve more passengers. (E[H] appears in the formula simply as a normalizing factor making the area under the density function equal 1.)

Consider passengers who arrive during headways of length h. Their waiting time distribution is uniform on the interval (0, h), with cumulative distribution function (CDF)

$$\mathrm{P}(W \le w \mid h) = \begin{Bmatrix} \dfrac{w}{h} & \text{if } h > w \\ 1 & \text{otherwise} \end{Bmatrix}$$

The CDF of waiting time is found by multiplying this conditional probability by $f_{H,pax}(h)$ and integrating:

$$F_W(w) = \mathrm{P}(W \le w) = \int_0^w \frac{h}{\mathrm{E}[H]} f_H(h)dh + \int_w^\infty \frac{w}{\mathrm{E}[H]} f_H(h)dh$$

$$= \int_0^w \frac{h}{\mathrm{E}[H]} f_H(h)dh + \frac{w}{\mathrm{E}[H]}[1 - F_H(w)]$$

Taking the derivative with respect to w yields the desired result:

$$f_W(w) = [1 - F_H(w)] / \mathrm{E}[H] \tag{5}$$

Very simply, the probability of waiting an amount of time w is proportional to the fraction of headways that are greater than w.

To illustrate, suppose headways have a uniform distribution on the interval [h_{min}, h_{max}]. That distribution and the resulting waiting time distribution are pictured in Figure 1. The waiting time distribution is constant from 0 to h_{min}, because waits of less than h_{min} can occur with any headway; it then falls linearly to a value of 0 at h_{max}, reflecting the fact that waits of length w can only occur with headways that are at least that long, which become rarer as w increases.

From this waiting time distribution, extreme values such as $W_{0.95}$ (the 95-percentile waiting time) can readily be calculated. While a closed form expression for $W_{0.95}$ can be obtained for the uniform headway distribution, it cannot be derived for a general headway distribution.

Constructing the Waiting Time Distribution from Headway Data

The distribution of waiting time can also be generated from a set of n observed headways. The waiting time distribution takes the form illustrated in Figure 2. Each headway h_j is represented by a rectangle, flush left, of length h_j and height $1/T$, where $T = \Sigma\, h_i$ is the total amount of time represented by the set of headways.

Corresponding calculations are shown in Table 1, columns *(a)*-*(e)*, yielding the cumulative distribution of waiting time $F_W(w)$. Headways are numbered in decreasing order, but listed in increasing order. Define $h_{n+1} = 0$, and $F_W(0) = 0$. Let $\Delta h_i = h_i - h_{i+1}$. Then beginning with $i = n$, apply the recursion

$$F_W(h_i) = F_W(h_{i+1}) + i\ \Delta h_i / T \qquad (6)$$

Between values of h and $F_W(h)$ found in the table, linear interpolation applies. For example, one can directly read that the fraction of passengers waiting 10 minutes or less is 93.7%; by interpolation, one can find that the 95-percentile waiting time is 10.6 minutes.

A convenient way of presenting the passenger waiting time distribution is to show the fraction of passengers in various waiting time ranges or "bins." Choosing bin thresholds corresponding to various levels of customer expectation allows one to see what fraction of passengers had various levels of service. Natural thresholds for a 3-level gradation ("good," "marginal," "poor") are $H_{schedule}$ and $H_{schedule}$+x; the lower threshold, because if headways are perfectly regular, no passenger's wait will exceed E[$H_{schedule}$]; the second, using a value of x (say, 2 minutes) reflecting customer expectations.

While bin frequencies can be calculated from the cumulative waiting time distribution, they can also be calculated directly, as shown in Table 1, columns *(f)*-*(i)*. Because each minute of a headway represents the passengers arriving during that minute, the minutes belonging to each headway are distributed into bins, filling the low waiting time bins first. For example, the 9-minute observed headway contributes 8 minutes' worth of passengers to the bin [0-8] and 1 to the bin [8-10]. In the last row, one can see that 2.1% of the passengers waited longer than 12 minutes, and 6.3% waited more than 10 minutes, *i.e.*, more that $H_{schedule}$+2 minutes.

Being able to determine the waiting time distribution supports service standards related to extreme values of waiting time, such as "No more than 5% of passengers should have to wait longer than $H_{schedule}$ + 2 minutes" which can be restated as

$$W_{0.95} \leq H_{schedule} + 2 \qquad (7)$$

A similar standard, except that it uses 90-percentile waiting time for bus and 98-percentile waiting time for metro, is published by AFNOR Certification (6) and used in such cities as Paris, Brussels, and Lyon to certify service quality.

Sensitivity to Service Reliability

Extreme values of waiting time are far more sensitive to headway variability than is E[W], as demonstrated in Table 2. There, for various levels of cv_H, mean, 90-, and 95-

percentile values of waiting time are shown, calculated using numerical integration and assuming normally distributed headways. The final column shows results for the data of Table 1. One can see, for instance, that an increase in cv_H from 0 to 0.35 causes $W_{0.95}$ to rise by 30%, while E[W] rises by only 12%.

Also shown for comparison in Table 2 are two related statistics: P[H > 10], a description of operational quality, and P[W > 10], a measure of service quality, both assuming E[H] = 8 minutes. Good news: it is much less likely for a passenger to have a long wait than for a bus trip to have a long headway – because only those passengers arriving in the early part of a long headway experience a long wait (unless there are pass-ups). Thus, for example, with normally distributed headways, a 95-percentile waiting time standard of 10 minutes (equation 7) can be met even if 24% of the headways exceed 10 minutes.

Budgeted and Potential Waiting Time

The 95-percentile waiting time can be interpreted as a measure of *budgeted waiting time*. When planning a trip, passengers must allow more time for waiting than mean waiting time; otherwise, they will reach their destination late about half the time! By budgeting an amount of time equal to $W_{0.95}$, one's probability of arriving late will be limited to 5%, meaning a weekday commuter will arrive late about once a month.

The difference between budgeted and mean waiting time can be called *potential waiting time*, $W_{potential}$. For short headway service,

$$W_{potential} = W_{0.95} - \mathrm{E}[W] \tag{8}$$

For example, a passenger going to work served by the headway distribution of Table 1 will budget $W_{0.95}$ = 10.6 min for waiting. Usually, he will wait less then this – on average, he will wait E[W] = 4.6 minutes. The balance, which varies from day to day and averages 6 minutes, is his potential waiting time, which he will spend waiting for work to begin.

As this example shows, potential waiting time is a hidden form of waiting, spent at the destination end of one's trip. Potential waiting time does not hurt passengers as much as platform time, because it is the part of the travel budget that is "redeemed," and can be spent more enjoyably or productively than waiting on the platform. Still, it represents a real cost to passengers, because it cannot be used as freely or reliably as if it were not encumbered by the travel budget. Our example traveler could use his 6 minutes of potential waiting time to stroll or start work early, but he could not use those 6 minutes to sleep later; neither could he rely on that time to get breakfast. When a journey involves a transfer from a short-headway route to a long-headway route, potential waiting time for the first leg of the trip will be usually be spent waiting at the transfer station.

Waiting Cost and Equivalent Waiting Time

A waiting cost function expresses the disutility of waiting in some standard units; we use equivalent minutes of in-vehicle time. Traditionally, waiting cost is based on nominal waiting time, and is assigned a unit cost between 2 and 2.5 minutes of in-vehicle time per minute of waiting. The premium placed on waiting time is understood to encompass both exposure to the waiting environment (weather, security, etc.) and "stress" related to the randomness of waiting time. The weakness of this approach is that it makes waiting cost completely insensitive to service reliability.

When cv_H is known or can be predicted, many researchers have based waiting cost on E[W], which accounts for service reliability through the term cv_H. However, this approach is incomplete. While E[W] captures the effect of service reliability on mean waiting time, it fails to account for passengers' need to budget more waiting time when service is less reliable.

We propose the following cost function for short headway waiting:

$$\text{waiting cost} = a\,\mathrm{E}[W] + bW_{potential} \tag{9}$$

where a and b are unit costs (minutes of in-vehicle time per minute of waiting).

It can be helpful to express waiting cost in a more natural unit, minutes of platform waiting, which we call *equivalent waiting time*:

$$W_{equivalent} = \mathrm{E}[W] + (b/a)\,W_{potential} \tag{10}$$

While original behavioral research to estimate coefficients a and b is beyond the scope of this project, it is possible to propose plausible values:

$$a = 1.5, \quad b = 0.75$$

Because platform waiting is time actually consumed, and involves exposure to the waiting environment, a should be greater than 1; and because potential waiting time is not actually consumed but only restricted in use, b should be less than 1. These values have been scaled so that with cv_H in the typical range [0.2, 0.35], they are consistent with the a unit cost of about 2.4 minutes of in-vehicle time per minute of nominal waiting.

One convenience of having the ratio b/a equal 0.5 is that equivalent waiting time becomes a simple average of mean and budgeted (95-percentile) waiting time:

$$W_{equivalent} = 0.5\left(\mathrm{E}[W] + W_{0.95}\right) \tag{11}$$

Reporting Measures and Formats

A graphical format for short headway waiting is shown in Figure 3's first three columns, using Table 1 data. (The last three columns pertain to long headway waiting.) Column 1 shows mean platform time, 95-percentile or budgeted waiting time, and equivalent waiting time. This format can be applied to single stop as well as to sets of stops (*e.g.*, a route or zone) by aggregating over stops, weighted by mean boardings at each stop.

The concept of "ideal" and "excess" waiting time, separating the impacts of planning from those of operation, is applied in columns 2 and 3 to budgeted and equivalent waiting time as well as to mean waiting time. "Ideal" is what would result from following the timetable; "excess" is the difference between actual and ideal. One can see that mean platform time is 0.6 minutes in excess of ideal (4.6 versus 4 min);

budgeted waiting time is 3.0 minutes in excess of ideal (10.6 versus 7.6 min); and equivalent waiting time is 1.8 minutes in excess of ideal (7.6 versus 5.8 min).

Because it accounts for the hidden waiting cost of potential waiting time, *excess equivalent waiting time* more accurately reflects the impact of operations on customer service than *excess mean waiting time*. If the company responsible for the Table 1 data were being judged on excess mean waiting time, they would be deemed responsible for 0.6 minutes per passenger, which is 15% greater than ideal. Using excess equivalent waiting time, they would be deemed responsible for imposing a cost of 1.8 minutes of waiting per passenger, a 31% increment above ideal.

LONG HEADWAY WAITING

With long headway service, passengers are expected to time their arrivals to meet a targeted scheduled departure. If service is punctual and early departures prohibited, platform waiting time can be quite small. One of the authors recalls a company bus that picked up his father and other employees from their village each day at 7 a.m. to take them to the factory. At 6:58 each morning, the stop was empty; at 7:01, it was empty again.

However, with long headway waiting, there are also forms of hidden waiting that contribute to passengers' waiting cost. We count four components of long headway waiting: platform waiting, potential waiting, schedule inconvenience, and synchronization cost, the latter two being unique to long headway waiting.

Offered Waiting Time

Determining passenger waiting time requires that some assumption be made about when passengers arrive at the station. One option is to assume that passengers arrive at a specified "arrival offset" (say, 0, 1, or 2 minutes) before the scheduled departure, perhaps consistent with advice published in a user's guide. The time until the bus departs can be called "offered waiting time." With long headway waiting, ridership on a given trip can be considered independent of schedule deviation; therefore, the distribution of offered waiting time corresponds directly to the distribution of schedule deviation. For example, if the arrival offset is set at 1 minute, the fraction of passengers with offered waiting time greater than 10 minutes is the fraction of trips departing more than 1 minute early or 9 minutes late.

With this framework, the distribution of schedule deviation is a good (and commonly used) measure of passenger service quality. As with short headway waiting, offered waiting time can be graded into various categories. The most common grading has just three categories, early, on-time, and late, with thresholds at 0 or –1 minute (the latter being an example of a non-zero passenger arrival offset) and 5 minutes.

To many customers, it is an exaggeration to call a bus "on time" when it is 4 or 5 minutes late. To put a sharper focus on service quality, one can use a finer grading for schedule deviations. However, because of the importance of extreme waiting times, it is perhaps more important to give attention to the fraction of passengers having large schedule deviations than to the fraction falling into a narrow "good service" category such as 0-3 minutes late.

Excess Platform Time

An alternative assumption about passenger behavior is that they base their arrival times on experience of service reliability. Let V = the schedule deviation of the trip a passenger is trying to meet, defined by

$$V = \text{departure time} - \text{scheduled departure time}$$

Early departures are represented by negative values of V.

We assume that passenger choose for their arrival time offset a low percentile value of V, one that limits the probability of missing their targeted bus. Taking that probability as 2%, we assume then that passengers arrive at or before $V_{0.02}$, the 2-percentile departure time, so that they will miss their targeted departure at most once every 50 trips.

Waiting time between when a passenger actually arrives and $V_{0.02}$ is hard to observe. However, it is largely unavoidable, having to do with uncertainty in access time. We want to focus on the part of waiting that is affected by service reliability. If service were perfectly reliable, there would be no waiting after $V_{0.02}$. The time that passengers wait after $V_{0.02}$ can therefore rightly be called *excess platform time*, with mean

$$W_{excess} = \mathrm{E}[V] - V_{0.02} \tag{12}$$

Reducing excess platform time is essentially a matter of reducing the early tail of the departure time distribution. Holding to a scheduled departure time is an obvious strategy for reducing excess platform time.

Potential Waiting Time

Just as with short headway service, passengers must be prepared to wait beyond the bus's expected departure time. Assuming that passengers budget for the 95-percentile departure time, expected potential waiting time is

$$W_{potential} = V_{0.95} - \mathrm{E}[V] \tag{13}$$

As with short headway service, potential waiting represents hidden waiting time spent at the destination end of a trip.

Just as excess platform time deals with the early tail of the departure distribution, potential waiting time deals with the late tail. Measures that shrink the late tail of the distribution, such as conditional priority at traffic signals (and to a lesser extent, holding, because it shifts the mean) will reduce potential waiting time.

Schedule Inconvenience

Schedule inconvenience, W_{SI}, is another hidden waiting time representing the difference between passengers' desired departure time and the departure time to which they are limited by the timetable. For example, suppose a passenger has to board a bus by 8:48 in order to get to work on time, and that scheduled departures around that time have scheduled, mean, and 95-percentile departure times, respectively, of {8:15, 8:17, 8:20} and {8:45, 8:47, 8:50}. IF she budgets for a 95-percentile departure time, she cannot use the 8:45 trip; she will have to take the 8:15 trip. Her average departure time is 8:17, and her potential waiting time is 3 minutes (because she can rely on boarding by 8:20). Her schedule inconvenience is then the interval between when she can count on boarding and her desired boarding time, or 28 minutes. She will "spend" her schedule inconvenience time at the destination end of her trip (for home-bound trip, however, W_{SI} will be spent at the origin side of the trip). Because she can rely on this time, she could use it for a planned activity such as breakfast or extended work time. W_{SI} is equally likely to take any value between 0 and h.

Schedule inconvenience differs from potential waiting time because a passenger can count on having the schedule inconvenience time each day, and can therefore plan activities to make better use of it. Therefore, its unit cost should be less than that of potential waiting time.

Synchronization Cost

Synchronization cost combines several burdens related to a passenger's need to adjust to a given schedule. One is the stress of arranging one's day to conform to a timetable. Most people would gladly suffer an extra minute of waiting for the freedom of arriving without the constraint of a bus or train schedule. Another is the waiting time between when a person actually arrives at a stop and their target arrival time, assumed to be $V_{0.02}$. The limits of human punctuality and randomness in access time demand that people arrive early, on average, if they don't want to miss their bus. Third is the stress of worrying about missing one's departure, which should increase with h. Fourth is wait of a full headway for the rare occasions for which if one misses the bus. These costs are either constant or proportional to headway; a reasonable proposal for synchronization cost, in equivalent minutes of in-vehicle time, is

$$\text{synchronization cost} = 2 + 0.05\,h \tag{14}$$

Waiting Cost

A passenger's cost of long headway waiting combines the four components just presented. For excess platform time and potential waiting time, proposed unit costs are a = 1.5 and b = 0.75, as for short headway waiting; a unit cost of 0.60 is proposed for schedule inconvenience. For an individual traveler with given schedule inconvenience, the long headway waiting cost function becomes

$$\text{cost} = 2 + 0.05h + 1.5(\mathrm{E}[V] - V_{0.02}) + 0.75(V_{0.95} - \mathrm{E}[V]) + 0.6W_{SI} \tag{15}$$

Averaging over all users at a stop,

$$\text{cost} = 2 + 0.35h + 1.5(\mathrm{E}[V] - V_{0.02}) + 0.75(V_{0.95} - \mathrm{E}[V]) \tag{15a}$$

The first two terms in equation 15a reflect the inevitable parts of waiting cost that arise simply from having a long headway; the last two terms are the *excess waiting cost* arising from unreliability. This cost function is illustrated in the long headway region of Figure 4 using example reliability cases described in the next section.

Example Results

Six reliability cases are used in Figure 4 and Table 3 to illustrate the sensitivity of waiting cost to reliability. Cases A-E represent worsening reliability, with normally distributed departure deviations with standard deviation σ_V ranging from 1.0 to 2.6 minutes. On-time performance (departures within a 5-minute window) for those cases ranges from 97% to 66%. Case D+OC represents a route with the same underlying randomness as case D, but with operational control (7) by means of holding, with stop-level scheduling tuned such that the earliest 30% of trips are held and are assumed to run, on average, 1 minute late.

As shown in the table, excess waiting cost varies from 4.4 to 11.2 minutes of in-vehicle time, showing strong sensitivity to reliability. It is especially instructive to compare cases D and D+OC. In terms of operational performance, operational control clearly offers an improvement, raising on-time performance from 73% to 92%. But what is the effect on passengers? Without a waiting cost function that accounts for service reliability, it is difficult to evaluate how much passengers benefit from this reliability improvement. With the proposed function, we have an answer: operational control reduces waiting cost by 4.5 minutes of equivalent in-vehicle time. Achieving the same benefit by reducing headway alone would require a headway reduction from 30 to 18 minutes, involving a 67% operating cost increase. This kind of analysis (it can also be seen in Figure 4) is needed to show the value of improving service reliability.

If excess waiting cost is expressed in units of equivalent minutes of *waiting time* rather than units of in-vehicle time, it becomes *excess equivalent waiting time*, defined just as with short headway waiting, but given for long headway waiting by

$$W_{ExcessEquiv} = \left\{ \mathrm{E}[V] - V_{0.02} \right\} + (b/a)\,(V_{0.95} - \mathrm{E}[V]) \tag{16}$$

Results for the six example cases are also given in Table 3. Analogous to short headway waiting, when the ratio b/a = 0.5, excess equivalent waiting time is simply the average of excess platform time ($\mathrm{E}[V] - V_{0.02}$) and excess budgeted waiting time ($V_{0.95} - V_{0.02}$).

The reporting format used for short headway waiting applies equally to long headway waiting. In Figure 3, the last three columns give results for cases A, D, and D+OC. One can readily see excess platform time, excess budgeted time, potential waiting time (the difference between the first two), and excess equivalent waiting time. This format makes it clear that the main difference between cases D and D+OC is that excess platform time was reduced – an expected consequence of holding, which primarily shrinks the early tail of the departure time distribution.

TRANSITION FROM SHORT- TO LONG-HEADWAY WAITING

The proposed short- and long-headway waiting cost models accomplish the objective of making waiting cost a function of service reliability. However, for planning and service design applications, one factor is still missing: the boundary at which passengers switch from short- to long-headway waiting. That boundary depends on service reliability, and so it should not be given exogenously; rather, it should be derived as a result of passengers choosing the least cost waiting strategy.

With historic data for a given route, the cost of the two waiting strategies can simply be evaluated and the best chosen. For planning applications, however, predicting waiting cost under the two strategies requires predicting extreme values of headway and of schedule deviation. Because headway deviation and schedule deviation are not independent, a service reliability model relating them is needed.

We take σ_V as the parameter describing a given reliability case. We assume a normal distribution of schedule deviations, with a correlation coefficient $\rho = -0.2$ between successive schedule deviations (implying a degree of bunching). With these assumptions, the headway distribution can be derived using the fact that a headway deviation is the difference between successive schedule deviations. We make one further assumption: that for $h < 10$ minutes, cv_H is maintained at the value achieved when $h = 10$ minutes (this prevents cv_H from rising to infinity as h becomes small). The headway cv corresponding to specified values of σ_V and h is then given by

$$cv_H(\sigma_V, h) = \sqrt{2-2\rho}\,\frac{\sigma_V}{\max(h, 10)} \tag{17}$$

A passenger's choice between long- and short-headway waiting depends on their schedule inconvenience, which varies across the population with a uniform distribution between 0 and h. Equating short headway waiting cost (equation 9) with long headway waiting cost (equation 15) for given h, one can determine an indifference value of schedule inconvenience, $SI^*(h)$. Passengers whose schedule inconvenience is lower than $SI^*(h)$ will prefer long headway waiting. When $SI^*(h) \leq 0$, everybody chooses short headway waiting; when $SI^*(h) \geq h$, everybody chooses long headway waiting.

A transition region exists where $0 < SI^*(h) < h$. In that region, the fraction of passengers choosing long headway waiting – the fraction with W_{SI} less than $SI^*(h)$ – is $SI^*(h)/h$. Their waiting cost uses the long headway cost function with $E[W_{SI}] = SI^*(h)/2$. Waiting cost for the remainder follows the short-headway cost function. Overall waiting cost is the sum of these short- and long-headway waiting costs, weighted by the fraction of passengers choosing either strategy.

"Indifference headway," for which half the passengers choose short- and half choose long-headway waiting, is shown in Table 3 for the six example cases, ranging from 8 to 14 minutes. Case C, a level of reliability that might be considered typical in large U.S. cities, has an indifference headway of 11 minutes, consistent with expectations. When operational control is applied to case D, indifference headway falls from 13 to 8 minutes – illustrating how with better operational control, passengers become more likely to rely on the timetable.

With the short- and long-headway waiting models thus harmonized, the waiting cost function for a full range of headways can be determined, as shown in Figure 4. At long and short extremes, waiting cost is linear with headway, with a steeper slope in the short headway region. At long headways, differences between reliability levels are substantial and constant. At short headways, differences between reliability levels are smaller. (As a practical matter, for very short headways, the principal benefit of improving regularity is balancing loads, which can both reduce operating cost and indirectly reduce waiting time by reducing the incidence of pass-ups.)

Clearly visible in Figure 4 is the smooth transition from the steeply sloped short-headway waiting cost function to the less steep long-headway waiting cost function. This smooth transition avoids the pitfall of the often-used bi-linear waiting cost model, in which all passengers shift from short- to long-headway waiting at a critical headway, resulting in a slope discontinuity known to cause bizarre results when applied in headway optimization (8).

PRACTICAL ISSUES

As we have shown, passenger waiting time is strongly affected by extreme values of headway and schedule deviation. Estimating extreme values requires a far greater sample size than does estimating a mean, making fleetwide AVL vital for estimating these measures. The need for a large sample size is most critical for estimating the 2-percentile schedule deviation. Using the "rule of 5," a sample of 5/0.02 = 250 observations is desirable – something only achievable with automatic data collection and some level of aggregation, such as combining several trips within a period of the day.

Passenger waiting times are based on vehicle *departure times* from a stop. Therefore, accuracy will be enhanced if AVL systems make this moment easy to determine – for example, by recording when the last door closes and the wheels start to roll, or when a vehicle leaves a GPS-defined zone around a stop. If an AVL system is configured to record only when a vehicle enters the zone around a stop, departure time estimates will be approximate.

At longer headways, respecting scheduled departure times helps passengers reduce their waiting time. However, getting the full benefit of this strategy requires publishing departure times for every stop, something common in European but not in U.S. practice, where departure times are typically published for timepoints only. This forces passengers boarding at intermediate points to estimate scheduled departure time conservatively, increasing their waiting time.

CONCLUSIONS

This paper shows that waiting cost has several components, of which two, platform and potential waiting, apply to both short- and long-headway waiting, and are strongly affected by service reliability. It develops waiting cost functions and measures of passenger waiting time that incorporate the effects of service reliability, allowing service reliability improvements to be translated into savings in passenger waiting time. The proposed models can help management measure service quality and evaluate investments that improve service reliability.

We have shown how with AVL data, one can calculate the distribution of passenger waiting time, as well as the extreme values of headway deviation and schedule deviation that are critical to passenger waiting cost.

For planning applications, we developed a waiting cost function that harmonizes short- and long-headway waiting strategies. By overcoming the discontinuity of the bi-linear cost model, it offers an improved framework for service design and demand forecasting.

ACKNOWLEDGEMENT

This research was funded in part by TCRP under project H28.

REFERENCES

1. Furth, P.G., Th.H.J. Muller, J.G. Strathman, and B. Hemily. *Uses of Archived AVL-APC Data to Improve Transit Management and Performance: Review and Potential.* TCRP Web Document 23 (Project H-28), 2003.
2. Kittleson & Associates *et. al. A Guidebook for Developing a Transit Performance-Measurement System.* TCRP report 88, 2003.
3. Kittleson & Associates *et. al. Transit Capacity and Quality of Service Manual*, 2nd ed. TCRP report 100, 2003.
4. Osuna, E.E. and G. F. Newell. Control Strategies for an Idealized Public Transportation System. *Transportation Science,* Vol. 6, 1972, pp. 52-72.
5. Wilson, N.H.M. *et. al.*, Service Quality Monitoring for High Frequency Transit Lines. *Transportation Research Record: Journal of the Transportation Research Board, No. 1349,* pp. 3-11, 1992.
6. AFNOR Certification. Service de Transport Urbain de Voyageurs: Reglement de Certification. Standard NF 286, 2002. http://www.afnor.fr, accessed May 1, 2005.
7. Muller, Th.H.J. and P.G. Furth. Integrating Bus Service Planning with Analysis, Operational Control, and Performance Monitoring. Proc., Intelligent Transportation Society of America Annual Meeting, Boston, 2000.
8. Jansson, K. Optimal Public Transport Price and Service Frequency. *J. Transport Economics and Policy* 27, pp. 33-50, 1993.

TABLES AND FIGURES

Tables

Figures

TABLE 1 Cumulative Waiting Time Distribution and Waiting Time Bin Frequencies

(a) h (min)	*(b)* Δh (min)	*(c)* i	*(d)* $i\,\Delta h$ (min)	*(e)* $F_W(h)$	*(f)* min in bin 0-8	*(g)* min in bin 8-10	*(h)* min in bin 10-12	*(i)* min in bin 12+
4	4	6	24	0.500	4			
5	1	5	5	0.604	5			
7	2	4	8	0.771	7			
9	2	3	6	0.896	8	1		
10	1	2	2	0.937	8	2		
13	3	1	3	1.000	8	2	2	1
T = 48			48		40	5	2	1
mean = 8					***83.3%***	***10.4%***	***4.2%***	***2.1%***

TABLE 2 Relationship of Headway Variation to Various Measures of Waiting Time

headway *cv*	**0**[1]	**0.15**[1]	**0.25**[1]	**0.35**[1]	**0.45**[1]	**0.38**[2]
E[wait] [3]	0.50	0.51	0.53	0.56	0.60	0.57
90-percentile wait [3]	0.90	0.93	0.99	1.08	1.18	1.14
95-percentile wait [3]	0.95	1.02	1.12	1.24	1.37	1.33
P[headway > 10 min] [4]	0	5%	16%	24%	29%	17%
P[wait > 10 min] [4]	0	0.3%	1.9%	4.6%	7.7%	6.3%

1. Headway distribution assumed normal.
2. Headway distribution as given in Table 1.
3. Relative to mean headway.
4. E[headway] assumed to equal 8.

TABLE 3 Results for Six Example Cases

Reliability Case	**A**	**B**	**C**	**D**	**E**	**D+OC**[1]
σ_V (min)	1.0	1.4	1.8	2.2	2.6	
Long Headway Results						
P[0 $\leq$ V $\leq$ 5]	97%	91%	82%	73%	66%	92%
E[V] – $V_{0.02}$ (min)	2.1	2.9	3.7	4.5	5.3	1.9
$V_{0.95}$ – E[V] (min)	1.6	2.3	3.0	3.6	4.3	2.9
excess waiting cost (min of in-vehicle time)	4.4	6.1	7.8	9.5	11.2	5.0
equivalent excess waiting time (min of waiting time)	2.9	4.1	5.2	6.3	7.5	3.4
***cv**$_H$* **at low headways**	0.15	0.22	0.28	0.34	0.40	0.24
indifference *h* (min)	7.9	9.4	11.0	12.7	14.4	8.0

1. Case D with operational control.

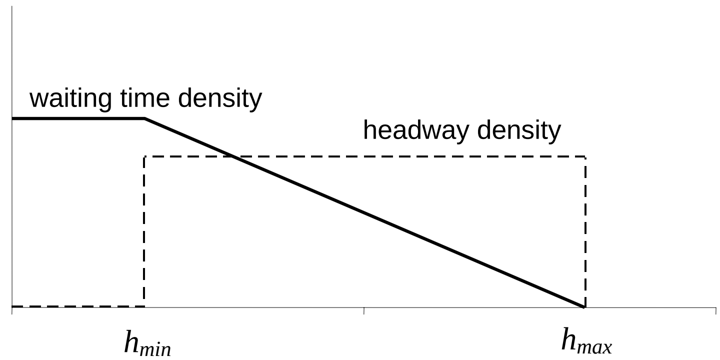

FIGURE 1 Uniform headway distribution and resulting waiting time distribution.

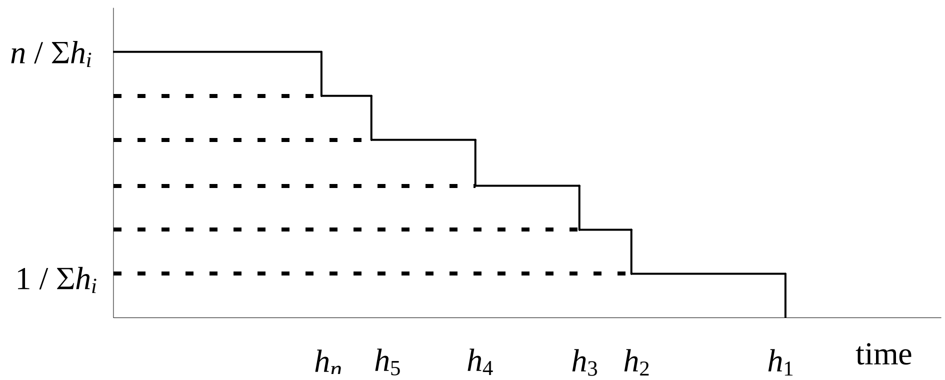

FIGURE 2 Waiting time density for an arbitrary set of headways.

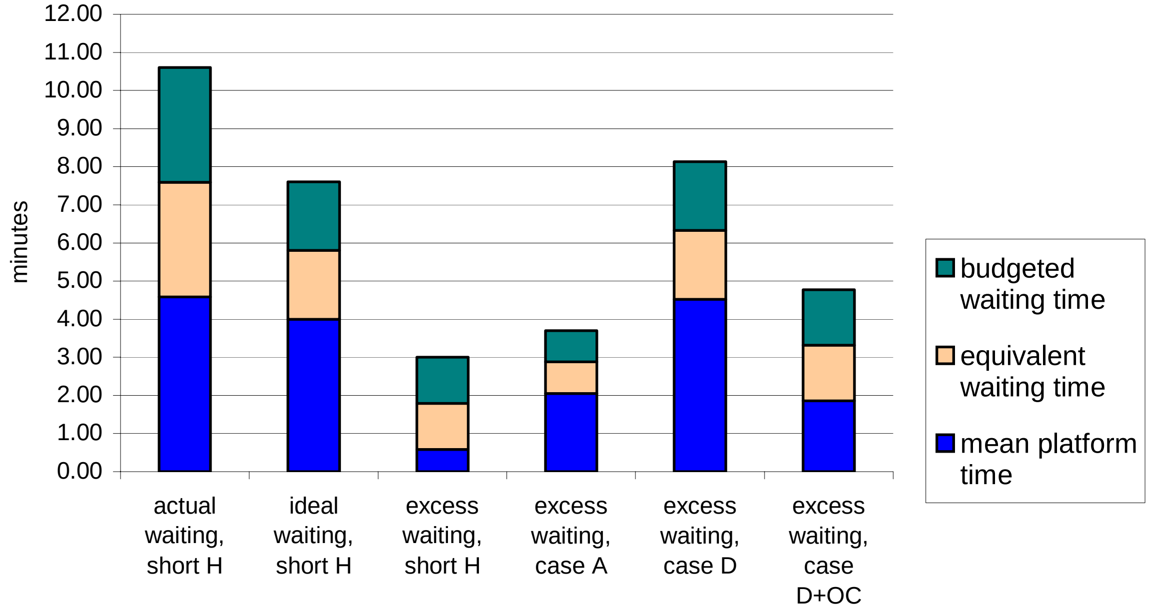

FIGURE 3 Passenger waiting time summary. Values are represented by cumulative heights.

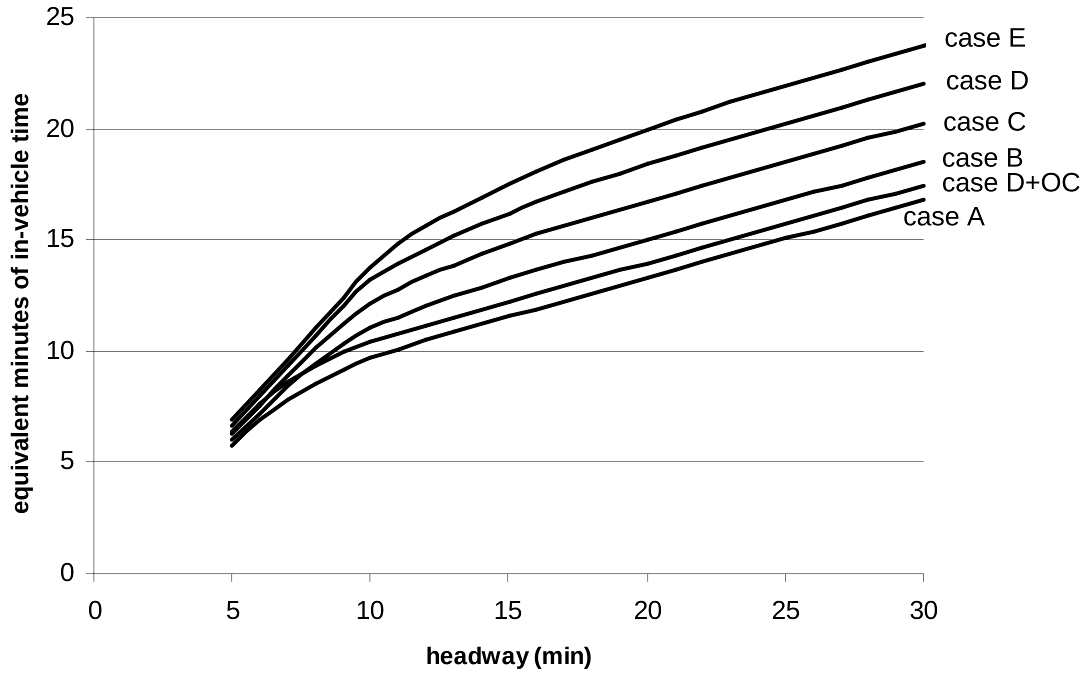

FIGURE 4 Waiting cost vs. headway for different cases of service reliability.