

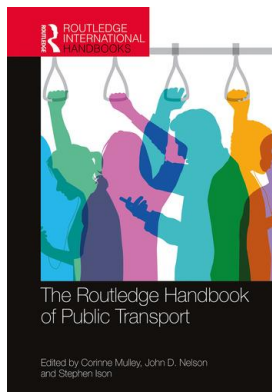
This article was downloaded by: 10.2.97.136

On: 22 Mar 2023

Access details: *subscription number*

Publisher: *Routledge*

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: 5 Howick Place, London SW1P 1WG, UK



## The Routledge Handbook of Public Transport

Corinne Mulley, John D. Nelson, Stephen Ison

### The use of smart card data to analyse railway station waiting times

Publication details

<https://test.routledgehandbooks.com/doi/10.4324/9780367816698-40>

Geoffrey Clifton

**Published online on: 13 May 2021**

**How to cite :-** Geoffrey Clifton. 13 May 2021, *The use of smart card data to analyse railway station waiting times from: The Routledge Handbook of Public Transport* Routledge

Accessed on: 22 Mar 2023

<https://test.routledgehandbooks.com/doi/10.4324/9780367816698-40>

**PLEASE SCROLL DOWN FOR DOCUMENT**

Full terms and conditions of use: <https://test.routledgehandbooks.com/legal-notices/terms>

This Document PDF may be used for research, teaching and private study purposes. Any substantial or systematic reproductions, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The publisher shall not be liable for an loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# 35

## THE USE OF SMART CARD DATA TO ANALYSE RAILWAY STATION WAITING TIMES

*Geoffrey Clifton*

### **Introduction**

In planning the public transport network, care needs to be taken to create a network that has the ability to encourage travel, and so trade-offs are made between coverage and frequency when the (subsidy) budget is constrained. The empirical evidence suggests that frequency is important in growing patronage (Currie & Wallis, 2008; Hensher et al., 2010). This underpins the best practice, outlined in Nielsen et al. (2005) and Walker (2008), which suggests that networks should concentrate resources on core routes because this gives a higher frequency for the same budget, which has a better chance of creating mode shift.

This best practice suggests that, beyond a certain headway, public transport services become frequent enough that customers will simply 'turn up and go'. The benefits of a 'turn up and go' service are discussed in Nielsen et al. (2005) and Walker (2008) and include the way this encourages greater usage of public transport by reducing the burden on customers of planning their journey, giving higher revenue for operators and a network that is simpler to plan (as transfers between frequent services no longer need to be timetabled) and market to potential customers. To properly plan requires some understanding of what 'turn up and go' means in the context of the urban area under consideration, as the actual frequency of a 'turn up and go' service may vary according to urban land patterns and typical journey lengths.

This chapter addresses the question of how frequent a service must be in order to trigger 'turn up and go' behaviour from passengers, using smart card data from the Australian city Sydney to examine how travellers' arrival times at rail stations react to different headways so as to identify the 'turn up and go' frequency. The chapter is structured as follows: The next section briefly sets out the literature context; second, the source data is discussed, followed by the methodology of the study. Finally, results of the analysis are presented, along with a discussion of the issues and limitations of the analysis.

### **Literature context**

Headway (the time between two consecutive train trip departures) is one of the train-network design factors which determines train patronage (Cipriani et al., 2012; Guihaire & Hao, 2008).

It is argued that different headways result in not only changes to the propensity of passengers to use a particular mode but also changes to passenger behaviour in terms of the arrival times of intending passengers at train stations relative to the scheduled departure time (Vuchic, 2017). Train passengers have been classified into random and non-random arrival time categories (Turnquist, 1981), also described as schedule-independent and schedule-dependent passengers (Lüthi et al., 2007; Vuchic, 2017). The wait time is also influenced by the headway regularity (Bowman & Turnquist, 1981; Turnquist & Bowman, 1980).

Chang and Hsu (2001) and Hsu (2010) modelled waiting time and distribution probability by scheduled and non-scheduled passengers of an intercity transit system and feeder bus services, respectively. These studies suggested that the waiting time of passengers taking intercity transit services significantly depends on the reliability of the feeder bus services. So, the waiting time is shorter than half of the headway with reliable feeder bus services and longer than the headways with unreliable bus services. In general, the quality of the feeder buses has a great impact on the waiting time even of scheduled passengers.

Many studies assume intending passenger arrivals follow a uniform distribution and therefore use half the headway as the median waiting time (see Chang and Hsu (2001); Marguier and Ceder (1984) and Welding (1957) for discussion of this), with other researchers reporting fixed values of waiting time for headways up to a given limit (e.g., 8.28 minutes for connecting services without capacity limitation reported by Hsu [2010]). Others have modelled arrival behaviour as a function of headways, including Jolliffe and Hutchinson (1975) who found that passenger arrivals at stations tend to be more random with more frequent services (shorter headways), with these random arrivals observed up to a service-frequency threshold. This fact has been confirmed in later studies on various urban transit systems (Turnquist, 1981; Vuchic, 2017). Mishalani et al. (2006) suggested that the frequency threshold for random-arrival passenger is 10 minutes for Columbus, US. Csikos and Currie (2008) found arrival behaviour also varied between peak (less random) and off-peak (more random) for the rail network in Melbourne, Australia, with headways between 10 and 20 minutes in both peak and off-peak.

Earlier researchers depended on direct observation of arrival times (Jolliffe & Hutchinson, 1975), but more recently smart card data has allowed for more passengers to be measured (Csikos & Currie, 2008; Pelletier et al., 2011). Various probability distributions for arrival times have also been applied, with Lüthi et al. (2007) estimating the median waiting time using headways based on a combination of a uniform distribution for schedule-independent passengers and a bounded distribution (Johnson's SB distribution, which is a transformation of the normal distribution that allows for skewness and kurtosis to be varied) for schedule-dependent passengers. This latter study found that passengers start to become schedule independent from a headway of 5 minutes but are more likely to time their arrivals to the schedule even for headways shorter than 5 minutes if they trust the timetable.

The literature context suggests that modelling arrival times at public transport stops using a uniform distribution is a reasonable assumption for short headway urban contexts, particularly at peak hours. However, the literature also suggests that, outside of that specific context, there are multiple factors that influence arrival times, primarily service headways but also time of day and purpose of journey. Furthermore, there is no consensus on which probability distributions could be used in place of the uniform distribution nor under which circumstances different distributions should be used. In particular, there is a need for more evidence on the headway at which intending passengers switch from schedule independence to schedule dependence.

## Data sources

A smart card ticketing system named Opal was introduced in 2013 for Sydney, Australia, replacing a paper ticketing system. This was based on similar technology to the Octopus card of Hong Kong and the Oyster card of London (Ellison et al., 2017).

Intending train passengers use their Opal to tap on upon arrival at the station, which allows station arrival times to be compared to the train timetable. Unfortunately, this same methodology cannot be applied to bus travellers, since bus services use onboard fare collection and the tap-on occurs once the service arrives.

Smart card data for one day (Tuesday 1 March 2016) has been provided by Transport for NSW, the state government department responsible for public transport. This data has been interrogated to look at the average time a passenger waits before boarding, classified by the average headway of services. Service headway is estimated from the GTFS timetable for the Sydney Trains network. Sydney Trains has a very complex network which serves approximately 1.5 million journeys each weekday. As of March 2016, the network included seven suburban train lines and four intercity lines, many of which sharing the same track and all of which use the same Opal card fare system.

## Methodology

Waiting times are calculated as the difference between the expected departure time from the GTFS timetable and the tap-on time from the Opal smart card data. In some cases, this leads to a waiting time greater than the headway, because in large stations, the tap-on location may be a distance from the platform from which the train departs. Under the assumption that passenger's behaviour is determined by the immediate future service frequencies, the average headway is estimated by averaging the time between consecutive services (i.e. headway) of the next three services after the departure time at which the passenger boards.

The Opal card data is structured by the time and location where passengers tap on linked to the time and location where passengers tap off to complete the journey. This journey may consist of one single movement by train, in which case calculating the waiting time is a straightforward matter of comparing the scheduled departure time of the train to the tap-on time of the intending passenger.

However, the complexity of the network means that approximately a quarter of rail passengers will need to transfer between trains to complete their journeys, and passengers may have a number of possible routes or itineraries they could take. Opal card data does not record these transfers, as Sydney Trains, as with many other train networks around the world, is a closed system where train passengers only leave traces at the tap-on and tap-off points. Passengers will only tap off and then tap on between segments of a train journey if they wish to leave the paid area of the station, and such journeys will be registered in the Opal system as separate records. This means it is necessary to determine which route the passenger took in order to determine their waiting time.

The Trip Planner data from Transport for New South Wales (TfNSW) provided a means to impute passengers' routings. Ho and Ho (2018) developed an application programming interface (API) script to query the trip planner website (<https://transportnsw.info/>) to download a mass dataset consisting of all possible route choices for trips between each of the origin-destination pairs observed in the Opal data. However, this dataset was only available for Tuesday 22 August 2017, whilst the Opal data was available for Tuesday 1 March 2016. This was not ideal

but still provides useful information, as there had been no major timetable revisions in the interval between March 2016 and August 2017 and both datasets were for a Tuesday where schools were in session, so travel patterns are assumed to be comparable.

As discussed in Ho and Ho (2018), the highest-probability route option for each passenger was selected after comparing the Opal card and trip planner datasets. This required calculating a generalised travel time for each of the possible routings that fitted within the tap-on and tap-off times recorded for the passenger.

The generalised travel time ( $gt_{in}$ ) for a particular route ( $i$ ) for a particular passenger ( $n$ ) is defined as:

$$gt_{in} = \text{interplatform time}_{in} + \text{waiting time}_{in} \times \alpha + \text{number of transfers}_{in} \times \beta + \gamma \times \text{egress time} \tag{Equation 35.1}$$

where:

$$\text{inter platform time} = T_{arrival} - T_{departure} \tag{Equation 35.2}$$

$$\text{waiting time} = T_{departure} - T_{tap-on} \tag{Equation 35.3}$$

$$\text{egress time} = T_{tap-off} - T_{arrival} \tag{Equation 35.4}$$

and the assumed parameters are  $\alpha = \gamma = 2.5$  and  $\beta = 5$  minutes.

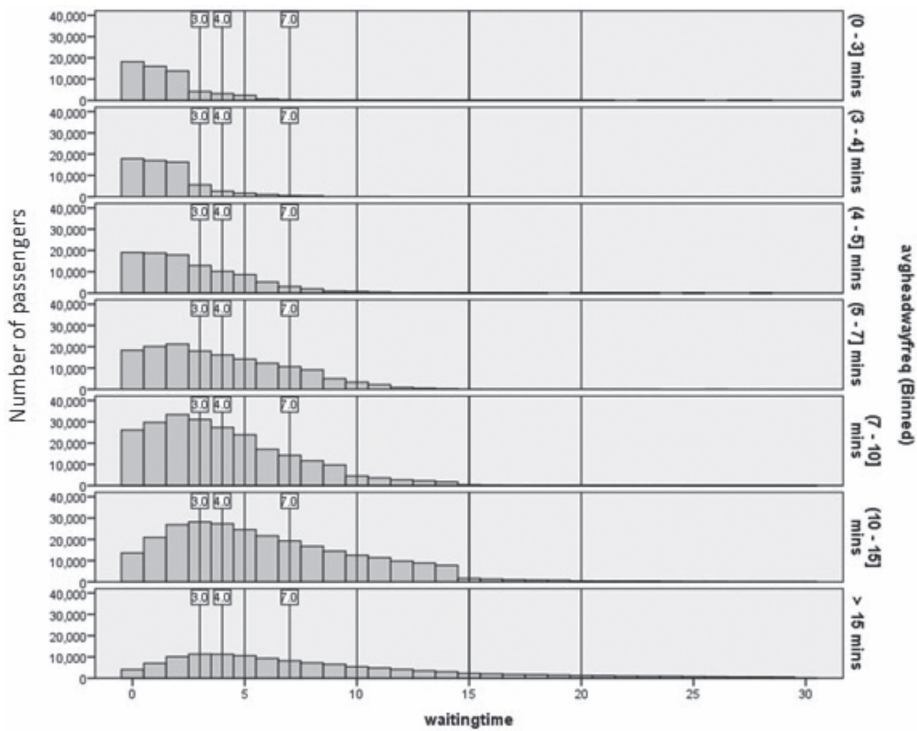


Figure 35.1 Distribution of wait time by service headway (time between two consecutive services) across all time periods

The probability of route  $i$  being selected by person  $n$  is calculated using the exponential of generalised time for each of the possible routes using the formula:

$$Probability_{in} = \exp(-0.3 \times gt_{in}) / \sum (\exp(-0.3 \times gt_n)) \tag{Equation 35.5}$$

with the sum being over all possible routes for passenger  $n$  and a co-efficient of  $-0.3$  based on the assumed travel time elasticity. Using this utility-maximisation heuristic, the probability of each itinerary  $i$  being selected by the train user  $n$  can be calculated and the itinerary with highest probability used to estimate the waiting time for each passenger (for details, see Ho and Ho [2018]).

For this chapter, the elapsed times between each of the four possible departure times were averaged to estimate the average headway around the time of each passenger’s arrival at the station. Average headways do not, however, distinguish between even and uneven headways, with an average of 4 minutes being the same for potential departure times at 08:04, 08:08, 08:12 and 08:16 given an even headway of 4 minutes and potential departure times of 08:04, 08:06, 08:14 and 08:16. Finally, waiting times were classified to different ranges of average headways to investigate the distribution of wait time by service headway (Figure 35.1).

### Analysis

If all passengers were schedule independent, then arrival times at stations would be random and waiting times would follow a uniform distribution. Schedule-dependent passengers would skew the distribution of waiting times away from the uniform with a cluster of arrivals around the scheduled train departure time, leading to an average waiting time of less than half the headway. If passenger arrival time behaviour is influenced by the headway, then we would expect the uniform distribution to fit observed arrival time behaviour for shorter headways, but we would also expect the uniform distribution to be rejected for longer headways.

For this analysis, only passengers with evenly distributed headways were used; these were either 3, 10, 15, 30 or 60 minutes. Next, the accumulated proportion of passengers who arrived within half the headway were calculated for each of the headways (Table 35.1). The median waiting times for each headway were also calculated (Table 35.2). To explore time of day effects, the data was divided into five times of day (ToDs): 07:00 and earlier (ToD1), 07:01 to 09:00 (ToD2), 09:01 to 16:00 (ToD3), 16:01 to 18:00 (ToD4) and 18:01 and later (ToD5).

Table 35.1 Percentage of people who arrived within half the headway segmented by time of day (rows) and by average headway between services (columns) and associated  $z$ -scores

Time of day (TOD)	3 minutes	10 minutes	15 minutes	30 minutes	60 minutes
ToD 1	51%	58%	72%	89%	n/a
$z$ -score	0.21	6.89	47.67	26.18	n/a
ToD 2	51%	61%	74%	85%	96%
$z$ -score	1.67	22.83	74.44	20.93	19.25
ToD 3	53%	53%	60%	71%	91%
$z$ -score	2.56	8.59	50.04	44.62	37.18
ToD 4	50%	52%	63%	78%	86%
$z$ -score	0.86	5.03	47.59	24.80	10.81
ToD 5	58%	50%	59%	77%	92%
$z$ -score	3.45	0.79	31.26	35.42	14.79

Table 35.2 Median waiting time of train passengers in Sydney segmented by time of day (rows) and by average headway between services (columns)

Time of day (TOD)	3 minutes	10 minutes	15 minutes	30 minutes	60 minutes
ToD 1	1	4	5	6	n/a
<i>z</i> -score	0.53	6.89	41.51	22.65	n/a
ToD 2	1	3	4	6	9
<i>z</i> -score	2.05	16.74	65.25	18.74	21.00
ToD 3	1	4	6	9	10
<i>z</i> -score	2.87	8.59	48.45	40.72	35.84
ToD 4	1	4	5	8	13
<i>z</i> -score	1.04	5.03	38.31	19.71	11.28
ToD 5	1	4	6	8	11
<i>z</i> -score	2.48	0.79	30.36	31.30	14.28

Contingency tables and chi-square/analysis of variance (ANOVA) tests were carried out to test the variance of waiting time distribution across each of the five times of day. Table 35.1 shows the percentage of passengers who arrived within half the average headway of the scheduled service segmented by time of day (rows) and by average headway (columns). Purely random arrivals would lead to 50% of passengers arriving within half the average headway, with the other 50% of passengers waiting more than half the average headway. Schedule-dependent passengers would be expected to time their arrival at the station to the departure time of the train. If all passengers were schedule dependent, then the percentage of people who arrived within half the headway would be expected to be close to or equal to 100%.

Consistent with the findings in other studies, the percentage of people who arrive within half the headway varies with the average headway. Over 75% of passengers arrive within half the headway for 30 minute and 60 minute headways, with the exception of interpeak passengers facing a headway for 30 minutes, where the proportion is only 71%. On the other hand, the proportion of people who wait more than half the headway is much higher for shorter headways, suggesting passengers arrive more randomly.

The actual distribution of waiting times can be tested against the uniform distribution using the *z*-scores shown in Table 35.1. These compare the observed proportion of passengers who arrive within half the headway to the proportion that would be expected to arrive within half the headway if passengers were schedule independent (i.e. 50%). The *z*-scores show that arrivals are consistent with the uniform distribution for 3-minute headways, but the uniform distribution is strongly rejected for headways of 10 minutes or greater, except for evenings (TOD5), where passenger behaviour is still consistent with schedule independence at headways of 10 minutes but not 15 minutes. This implies differences in arrival time behaviour by time of day; however, these differences do not seem to translate to differences in the median waiting times shown in Table 35.2.

The *z*-scores in Table 35.2 compare the observed median waiting time against the median waiting time that would be expected under the uniform distribution, which is half the waiting time. These are consistent with the results shown in Table 35.1.

The *z*-scores in both tables need to be placed in context. The results suggest that the arrival times of *all* Sydney train passengers can be modelled as ‘turn up and go’ where the headway is less than 10 minutes (less than 15 minutes in the evening). However, at a headway of 10 minutes, there is still between 39% (during the morning peak TOD2) and 50% (evenings TOD5)



of passengers who wait for more than half the headway. This suggests that the ‘turn up and go’ standard headway of between 10 and 15 minutes used in most of Australia is still appropriate for a large proportion of public transport users. However, it is possible that train passengers behave differently from travellers by other modes, and there may also be differences between cities.

### **Conclusions and recommendations**

It is clear that there is no single ‘turn up and go’ headway that applies to all passengers, with some passengers still exhibiting ‘turn up and go’ behaviour at headways longer than 15 minutes and others seeming to be schedule dependent even at headways below 10 minutes. In line with the findings of Csikos and Currie (2008), this behaviour appears to have some variation with time of day. The next step would be to model the aggregate arrival times as a mix between a uniform distribution and a non-uniform distribution reflecting the differing proportions of schedule-independent and schedule-dependent passengers at different times of day under different average headways.

Walking time from the point where passengers tap on to the departure platform varies with the size of the station and is a potential limitation of the analysis. Smaller stations tend to have the Opal card reader on the platform itself, whilst larger stations have gate lines that may be some distance away from the departure platform. This was considered a potential issue for the analysis, so actual walk times were measured empirically for the larger train stations (Central, Redfern, Town Hall, Wynyard, Parramatta, Strathfield). However, most of these walk times were actually 1 minute or less, and the tap on and tap off data was only recorded to the nearest minute. Therefore, variances in the walking time from entrances were excluded in the calculations.

A further limitation is the inability to distinguish between passengers who arrive at train stations by feeder bus versus passengers who arrive by other means. This is a potential issue because feeder bus passengers cannot control their arrival time at the station to the same extent as passengers who walk to the station. The final limitation of the analysis is the inability to transfer this analysis to the bus network given that arrivals at bus stops are not recorded by the Opal system.

There is much further work to be done on the long-term impact of COVID-19 on travel behaviour. This chapter divided public transport users into schedule-dependent and schedule-independent cohorts, and it will be worth examining the long-term impacts of COVID-19 on the travel behaviour of these two groups. It may well be the case that fear of exposure to infection will lead rail travellers to minimise their waiting time at railway stations. This would imply a switch to more schedule-dependent behaviour if passengers match their arrival at their departure station to the timetable and/or real-time information to minimise waiting time on the platform. Furthermore, a long-term growth in working from home, coupled with more in-home leisure activities and online retail usage would reduce the demand for public transport, and this could well affect one of the cohorts more than the other. It would be worth conducting the analysis undertaken in this chapter again to examine the extent to which any of these changes occur as travellers adjust to the ‘new normal’.

### **Acknowledgements**

Data for this project was supplied by Transport for NSW (TfNSW) and analysed by (Thi Kim) Loan Ho according to an API devised by Chinh Ho. An earlier draft of this chapter was presented at the 14th International Conference on Advanced Systems in Public Transport (CASPT) and the 4th International Workshop and Symposium on Research and Applications



on the Use of Passive Data from Public Transport (Transit Data), Brisbane, Australia, 23 to 25 July 2018. The authorship of that paper was Geoffrey Clifton, Corinne Mulley, Thi Kim Loan Ho, Chinh Ho and Barbara Yen.

## References

- Bowman, L. A., & Turnquist, M. A. (1981). Service frequency, schedule reliability and passenger wait times at transit stops. *Transportation Research*, 15, 465–471.
- Chang, S. K. J., & Hsu, C. L. (2001). Modeling passenger waiting time for intermodal transit stations. *Transit Planning, Intermodal Facilities, and Marketing*, 1753, 69–75.
- Cipriani, E., Gori, S., & Petrelli, M. (2012, February). Transit network design: A procedure and an application to a large urban area. *Transportation Research Part C: Emerging Technologies*, 20(1), 3–14. doi:10.1016/j.trc.2010.09.003
- Csikos, D., & Currie, G. (2008). Investigating consistency in transit passenger arrivals – Insights from longitudinal automated fare collection data. *Transportation Research Record*, 2042, 12–19. doi:10.3141/2042-02
- Currie, G., & Wallis, I. (2008). Effective ways to grow urban bus markets – A synthesis of evidence. *Journal of Transport Geography*, 16(6), 419–429. doi:10.1016/j.jtrangeo.2008.04.007
- Ellison, R. B., Ellison, A. B., Greaves, S. P., & Sampaio, B. (2017). Electronic ticketing systems as a mechanism for travel behaviour change? Evidence from Sydney's opal card. *Transportation Research Part A: Policy and Practice*, 99, 80–93. doi:10.1016/j.tra.2017.03.004
- Guihaire, V., & Hao, J. K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, 42(10), 1251–1273. doi:10.1016/j.tra.2008.03.011
- Hensher, D. A., Mulley, C., & Yahya, N. (2010, March). Passenger experience with quality-enhanced bus service: The Tyne and Wear 'superoute' services. *Transportation*, 37(2), 239–256. doi:10.1007/s11116-009-9240-x
- Ho, Q. C., & Ho, L. (2018, July 23–25). *Crowding on trains and platforms: A new approach to big data*. Conference on Advanced Systems in Public Transport, Brisbane, Queensland, Australia.
- Hsu, S. C. (2010). Determinants of passenger transfer waiting time at multi-modal connecting stations. *Transportation Research Part E: Logistics and Transportation Review*, 46(3), 404–413. doi:10.1016/j.tre.2009.12.002
- Jolliffe, J. K., & Hutchinson, T. P. (1975). A behavioural explanation of the association between bus and passenger arrivals at a bus stop. *Transportation Science*, 9(3), 248–282. doi:10.1287/trsc.9.3.248
- Lüthi, M., Weidmann, U., & Nash, A. (2007). *Passenger arrival rates at public transport stations*. 86th Annual Meeting of the Transportation Research Board (TRB).
- Marguier, P. H. J., & Ceder, A. (1984). Passenger waiting strategies for overlapping bus routes. *Transportation Science*, 18(3), 207–230. doi:10.1287/trsc.18.3.207
- Mishalani, R. G., McCord, M. M., & Wirtz, J. (2006). Passenger wait time perceptions at bus stops: Empirical results and impact on evaluating real-time bus arrival information. *Journal of Public Transportation*, 9(2), 18. www.nctr.usf.edu/wp-content/uploads/2010/03/JPT-9-2-Mishalani.pdf
- Nielsen, G., Nelson, J. D., Mulley, C., Tegnér, G., Lind, G., & Lange, T. (2005). *HiTrans best practice guide 2: Public transport – Planning the networks*. Civitas Consultants.
- Pelletier, M. P., Trepanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557–568. doi:10.1016/j.trc.2010.12.003
- Turnquist, M. A. (1981). Strategies for improving reliability of bus transit service. *Transportation Research Record*, 818, 7–13. https://trid.trb.org/view/174188
- Turnquist, M. A., & Bowman, L. A. (1980). The effects of network structure on reliability of transit service. *Transportation Research Part B: Methodological*, 14(1–2), 79–86. doi:10.1016/0191-2615(80)90034-X
- Vuchic, V. R. (2017). *Urban transit: Operations, planning, and economics*. John Wiley & Sons.
- Walker, J. (2008). Purpose-driven public transport: Creating a clear conversation about public transport goals. *Journal of Transport Geography*, 16(6), 436–442. doi:10.1016/j.jtrangeo.2008.06.005
- Welding, P. I. (1957). The instability of a close-interval service. *Operational Research Quarterly*, 8(3), 133–148. doi:10.2307/3007157