A Passenger-Focused Management Approach to the Measurement of Train Delay Impacts

Paper submitted for the Transportation Research Board 92nd Annual Meeting (2013) and for consideration for the Transportation Research Record

Date submitted: 1st August 2012
Date revised: 15th November 2012

Corresponding author: Alexander Barron
Railway and Transport Strategy Centre, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
E-mail: alexander.barron@imperial.ac.uk

Patricia C. Melo
Railway and Transport Strategy Centre, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
E-mail: patricia.melo@imperial.ac.uk

Judith M. Cohen
Railway and Transport Strategy Centre, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
E-mail: judith.cohen10@imperial.ac.uk

Richard J. Anderson
Railway and Transport Strategy Centre, Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London
E-mail: richard.anderson@imperial.ac.uk

WORD COUNT:
Abstract: 225 words
Text: 4824 words
Tables: (1) 250 words
Figures: (7) 1750 words
References: 443 words
Total: 7492

ABSTRACT

Train delay incidents have major impacts on transit service reliability and on customer satisfaction. Operators have long focused efforts on preventing incidents from occurring. While this is important, the fact that all transit operations inevitably face some degree of delays and disruptions due to incidents means that operators must also dedicate attention to reducing the duration of incidents and reducing the time it takes to restore normal operations after incidents occur. To be able to do this it is necessary to measure the total impact of incidents on trains and on customers. This research uses data from the CoMET and Nova metro benchmarking groups to investigate the ways in which transit operators can better measure the full impacts of incidents – both on train service and on customers – leading to a passenger-focused approach. The key benefit of a passenger-oriented approach is that it enables transit managers to better direct resources for incident response and recovery, as well as support the case for strategic investments. This research has showed that most operators only measure and report the frequency of incidents. Of the 22 metros interviewed, only two were able to provide detailed data to estimate the number of passengers affected by incidents. It is no coincidence that the only two metros able to provide detailed data are in fact two of the most reliable in the group.

KEYWORDS: Train delay incidents, reliability, CoMET and Nova metros

1. INTRODUCTION

All transit[1] operators face delays and disruptions due to incidents. This includes incidents more directly within the operator's control, such as equipment failures and train or staff availability, as well as those less within the operator's control, like suicides or power supply problems. These delays due to incidents not only affect the reliability of the transit service but also have a direct link to customer satisfaction. Furthermore, major delays are the subject of media attention in most countries, and delays can impact both an operator's public perception and its relationship with the government, potentially affecting decisions on funding.

The frequency of such incidents varies significantly around the world. Based on available data from the CoMET[2] and Nova[3] metro benchmarking groups, incident frequency ranges from about once per week for very small and/or very reliable metros up to approximately 50 or 100 times per day for very large and/or very unreliable metros. Furthermore, the impacts of these incidents are large – each, on average, is estimated to cause nearly 30 minutes of train delay and more than 300 hours of passenger delay.

The primary focus for transit operators has long been to try to prevent such incidents from occurring – and many operators have significantly reduced the frequency of incidents over time. However, it is also critical that operators dedicate attention to reducing the duration and impact of incidents (such as the time that the service remains stopped) and reducing the time it takes to restore normal operation after incidents occur. This is especially true for operators with capacity constraints, because the consequences of any incident will be great, as well as for those operators who have already reduced the frequency of incidents, because the remaining incidents will more likely be larger incidents with greater impacts.

The first step to doing this – to improving response to and recovery from incidents causing delays – is to fully understand the impacts of the incidents. Performance measures in this area of service reliability typically focus only on the frequency at which incidents or delays occur, such as with the mean distance between failures (MDBF). However, the impacts of incidents are more important than the number. To truly understand reliability, performance measures that capture the total impact of incidents to trains and ultimately to customers are also needed, such as train hours of delay and passenger hours of delay.

This is important because the impacts of incidents increase with the duration of incidents and the time they occur (peak vs. off-peak). Figure 1 illustrates this fact. Although we do not have sufficient data to test the actual shape of the relation between the impact of incidents and their duration, the data suggest that the relation is steeper during peak period. Incidents that occur during peak times or at busy locations have much greater impacts than those at the outer end of a long metro line late on a Sunday night. The red arrows in Figure 1 show that if an incident occurs on a line with a higher train frequency (where there are more trains) or at peak time (when more

---

[1] While this paper uses examples from metros, the same principles can be applied to all transit operations.

[2] CoMET: The metros in Beijing, Berlin, Guangzhou, Hong Kong, London, Mexico City, Madrid, Moscow, New York, Paris, Santiago, Shanghai, São Paulo and Taipei.

[3] Nova: The metros in Bangkok, Barcelona, Buenos Aires, Brussels, Delhi, Lisbon, Milan, Montréal, Naples, Newcastle, Rio de Janeiro, Singapore, Toronto and Sydney.

people are on those trains), the impact of that incident increases even more quickly. On the other hand, the green arrows show that, if the incident occurs in a less busy location or on a long line with alternative branches that can be used instead, that exponential growth in impact will be less than it might otherwise have been.

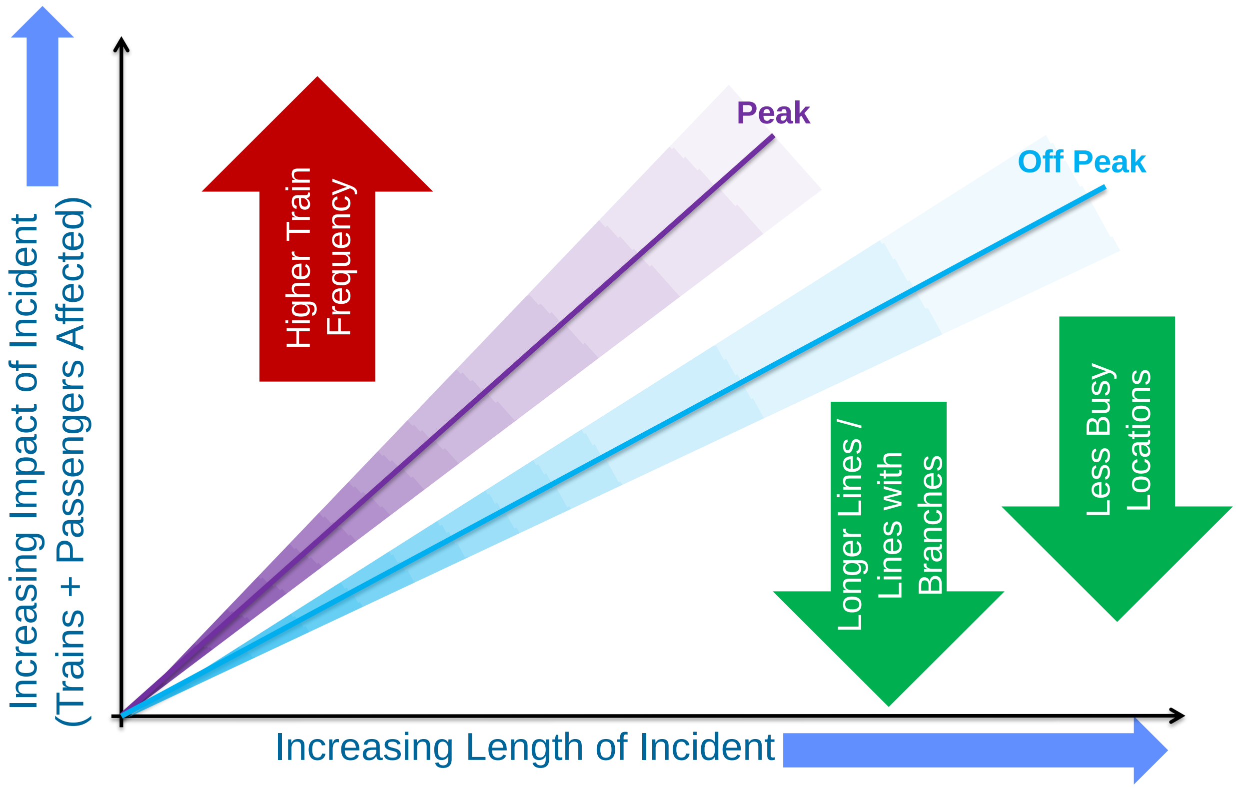

FIGURE 1 The impact of incidents increases exponentially with duration

Therefore, this research investigates how transit operators can improve the way they measure and understand the full impacts of incidents. By doing so, transit managers can better direct resources to reduce these impacts, both in the short-term (with staffing, equipment, and procedures) and the long-term (by strategically investing in improvements).

Data about incidents and delays on metros, and how they are measured, were collected through questionnaires and interviews designed to fit the specific objectives of this research. The participating metros were asked to provide detailed data to understand the total impacts of incidents. The relative severity of the different types of incidents was then compared using a number of different performance indicators.

Of the 22 metros that responded to the questionnaire, all were able to provide the most basic data such as the number of incidents and mean distance between failures, and nearly all metros were able to provide incident data disaggregated by cause. Far fewer metros were able to provide detailed incident data such as the duration of degraded or stopped service, and the impact on trains and passengers. Only two metros were able to provide sufficiently detailed incident data usable for an extended analysis of the total impacts of incidents. The aim of this extended analysis was to test whether passenger-focused reliability indicators would lead to different conclusions from those suggested by incident frequency indicators. Our main hypothesis was that the use of indicators focused only on the frequency at which incidents or delays occur can be misleading in providing information to transit managers about how to effectively manage the impact of incidents on passengers.

The analysis conducted in this paper has showed that transit operators need to collect and analyze a wide range of data to understand the impact of incidents and train delays on passengers. The evidence from our extended analysis indicates that there are important differences between the frequency of incidents, duration of service delay, and total disruption to trains and passengers by cause. As a result, achieving a reduction in the frequency of incidents of a certain type (e.g. rolling stock) may not deliver the largest benefit to passenger service reliability and customer satisfaction. The development of passenger-focused indicators of service reliability enables transit managers to make more cost-effective decisions about when and where to direct resources. This can be done by incorporating reliability into the economic appraisal of alternative strategies of incident prevention, response, and recovery.

The structure of the paper is as follows. Section 2 provides a discussion of the relevant literature on train service reliability. Section 3 describes the data available and outlines the methodology used. Section 4 presents and discusses the main results, while section 5 draws together the main conclusions.

2. RELATED LITERATURE

2.1 Monitoring train service reliability

Service reliability is a key success factor for mass transit systems and is often identified by passengers as the most important aspect determining service quality [e.g. 1, 2]. The concept of service reliability can be used to mean different, though related, things. In particular, service reliability is generally used in respect to service unpunctuality and delay. However, as discussed by Bates et al. [3, 4], the term reliability accounts both for the predictability of a given travel time being experienced by a passenger and the degree of variation around the average travel time [3-5]. In other words, service reliability depends both on the variability and predictability of travel times [6].

The first step in improving service reliability is to develop appropriate indicators. There is evidence showing that an increasing number of mass transit operators around the world are using indicators to monitor service reliability [7]. When considering the different indicators used to measure service reliability, however, a dichotomy between operator-oriented and passenger-oriented perspectives becomes evident. It is important to distinguish between the two perspectives because they often convey different perceptions of the degree of service reliability of the mass transit system.

Operator-oriented measures tend to focus on the vulnerability of the network to disruption and the operating performance of the network as compared to some agreed level of service (e.g. number of train cancellations, number of incidents, average punctuality). These measures tend to provide an aggregate view of the network's performance and, as a result, generally fail to reflect the user's experience [e.g 8, 9]. In contract, passenger-oriented indicators focus on the users' experience. Passengers are mostly concerned with the variability and uncertainty of the travel times of their journeys, not the average network performance. Typically, operator-oriented indicators include measures of service availability (e.g. number of train cancellations), average punctuality or regularity (e.g. percentage of trains arriving on-time), network vulnerability (number of incidents by cause), and total and average train delays (e.g. train hours delay). On the other extreme of the spectrum of indicators of train service reliability, we have passenger-oriented measures that describe the degree of travel time variability and uncertainty. Some of the most common indicators include the *travel time index*, *planning time index*, and *buffer time index*. The travel time index measures how much longer travel times are during peak compared to off-peak. The planning time index and the buffer time index measure the total time and the extra time, respectively, travellers should allow to ensure on-time arrival [10].[4]

Although these passenger-oriented indicators provide a description of train service reliability from the passenger's point of view, they are not necessarily useful to transit operators when making decisions about how and where to invest in order to enhance train service reliability. The main issue here is to develop indicators that reflect the impact of network failures on its users while also enabling transit operators to allocate resources to response and recovery strategies in a more effective way. Some of the reliability indicators discussed and proposed in this paper (see sections 3 and 4) fulfil this purpose. They can be interpreted as passenger-focused management-oriented indicators of train service reliability, and provide an intermediate level of measurement between the operator-oriented and passenger-oriented spectrum of train service reliability measures.

For many transit operators, moving from an operator-oriented perspective to a passenger-focused management-oriented perspective can be achieved simply by making a more comprehensive use of the data already available such as from signalling systems, train control systems, or automated fare collection systems (e.g. turnstiles/fare gates). The path to achieve this requires adjusting the use of existing data in order to improve conventional measures and/or develop new measures.

2.2 Incorporating train service reliability into cost-benefit analysis (CBA)

Besides providing a better representation of service quality as perceived by passengers, reliability indicators can also be used to inform better decision making when combined with cost-benefit (CBA) analysis. The latter allows the comparison and ranking of investments according to their expected outcomes in terms of improvements to travel time reliability. The failure to account for reliability in cost-benefit analysis results in the underestimation of an important part of the benefits of mass transit improvements.

Only a small number of countries include monetized values of reliability in the CBA of transportation investments [7, 11]. These include, to mention a few, Australia, Japan, Norway,

[4] Many indicators have been developed for road transportation based on either estimates or direct measurement of travel times, but are lacking for rail. For a full description of commonly used measures of travel time reliability see Lomax et al. [10] and Van Lint et al. [6].

New Zealand, the Netherlands, Sweden, and the United Kingdom. The incorporation of reliability into CBA requires the estimation of (i) the expected change in travel time reliability associated with a given investment, and (ii) knowledge of the monetary values of travel time reliability. Although an increasing number of studies provide estimates of the monetary value of travel time reliability, the current treatment of reliability in CBA is generally constrained by the lack of trustworthy parameters.

There are several review studies of existing monetary values of travel time reliability, including Bates et al. [4, 12], De Jong et al. [13], Preston et al. [14], Vincent and Hamilton [15], and Tseng et al. [16]. In general, the empirical evidence shows that passengers attribute a higher value to travel time reliability than to travel time. This means that passengers prefer to trade-off some additional journey time by a reduction in the variability around journey times. The two main parameters used to quantify this relationship are the *reliability multiplier* and the *reliability ratio*. The former is a measure of the value of the mean delay time and is expressed as the value of delay divided by the value of scheduled journey time. The second parameter consists of the ratio of the value of the standard deviation of journey time over the value of journey time. In a review of the empirical literature conducted by Preston et al. [14] it was shown that one minute of delay is valued as 1.25-3 times the value of one minute of journey time, and one minute of standard deviation of travel time is valued between 1 and 2.8 times more than one minute of journey time.

Delay elasticities can be used by transit operators to estimate the impact of service (un)reliability on patronage and revenue. Preston et al. [14] summarise existing evidence for the United Kingdom on the elasticity of rail journeys with respect to delay time. The short-run elasticity values reviewed suggest that an increase of 10% in delay time is associated with a decrease in patronage of 1.12%, 0.79%, and 3.63% for non-London short distance journeys, journeys in London and the Southeast, and journeys to/from airports respectively. This reinforces the importance, as discussed in Section 2.1, of monitoring train service reliability using passenger-focused indicators. Although system failures are generally the cause of rail service unreliability, reliability should be measured in terms of its consequences to travellers.

3. DATA AND METHODOLOGY

This section describes the data and research methods used to analyse metros' approaches to the measurement of train delays. The research was carried out by the Railway and Transport Strategy Centre (RTSC) as part of the CoMET and Nova metro benchmarking groups. CoMET and Nova are benchmarking consortia comprising 14 of the world's largest metro systems and 14 small to medium metros, respectively. Information shared by the metros is covered by an existing confidentiality agreement which requires results to be published in an anonymised form. Therefore metros will not be referred to by their names throughout this paper.

Relevant data were collected through questionnaires and interviews designed to fit the specific objectives of this research. The participating metros were asked to provide detailed data to understand the total impacts of incidents, as shown in Figure 2 below. Detailed definitions were provided for all data items requested. The main components include the items below, and were requested for each individual incident[5]:

1. **Date and time** at which the incident started,
2. **Incident cause** (see below),
3. **Duration of degraded service** (service slowed during an incident, or due to congestion following incident resolution),
4. **Duration of stopped service** (service completely stopped),
5. **Number of trains affected** (the initial train affected and any others cancelled or late as a result of the incident),
6. **Train delay time** (the sum of all delay time caused to all trains affected by the incident),
7. **Number of passengers affected** (the total number of passengers estimated to have been affected by the incident, including those on the train with a problem and those on subsequent trains that were delayed), and
8. **Passenger delay time** (the sum of all delay time caused to all passengers affected by the incident).

'Incident cause' refers to a list of causes developed by the RTSC over more than ten years as part of an annual performance benchmarking process. These definitions classify incidents into groups according to common causes as follows:

- Rolling Stock Related Incidents (e.g. doors),
- Signalling Related Incidents (e.g. interlockings failure),
- Other Equipment Related Incidents (e.g. power),
- Staff Related Incidents (e.g. driver fault),
- Passenger Related Incidents (e.g. suicide),
- No Fault Found Incidents, and
- Other Incidents (e.g. capital reconstruction, police/security).

The responses received were analysed by the RTSC first to identify missing values and potential reporting errors. Structured follow-up questions and telephone interviews were also conducted to ensure that the information received was clear, reliable and comparable.

The relative severity of the different types of incidents was compared, using a number of different performance indicators (see Figure 2 and Table 1). For this analysis, the incident types were grouped according to the reporting categories defined above. 'No fault found' and 'other' incidents were excluded because their causes were too diverse to be grouped together for analysis in this way. The high frequency of rolling stock and signalling incidents enabled 'total equipment related incidents' to be partly disaggregated into rolling stock, signalling, and all other equipment related incidents.

The following section presents and discusses some of the main results obtained from the analysis of the data received from the participant metros.

4. RESULTS AND DISCUSSION

[5] Incidents are defined as any incident, due to any cause, resulting in a delay to the train service of 5 minutes or more.

4.1 Overview of metros' approaches to the monitoring of incidents

Figure 2 gives an overview of the incident data that the 22 responding metros were able to provide. Of the metros that responded to the questionnaire, all were able to provide the most basic data such as the number of incidents and mean distance between failures, and 21 were able to provide the incident cause. The ability to disaggregate incident data by cause is fundamental to managing incidents and their impacts, because it underpins any effort to address common causes.

Twelve metros were able to provide the incident date, of which eleven could also provide the incident start time. Date and start time can provide useful information about the temporal patterns of incidents with different causes, and how those relate to service levels and passenger crowding. Some incident causes are more directly associated with service frequency and crowding levels than others; for example, equipment failures may occur randomly or be driven by overcrowding or intense service levels. In addition, incident date and time are needed to estimate passenger crowding levels, which in turn are needed to estimate impacts on passengers.

Few metros were able to provide detailed incident data such as the duration of degraded or stopped service, and the impact on trains and passengers. Furthermore, those metros that were able to provide some of these data items were generally not able to provide all of them. Only two metros among the 22 respondents were able to provide sufficient incident data usable for our detailed analysis of the total impacts of incidents.

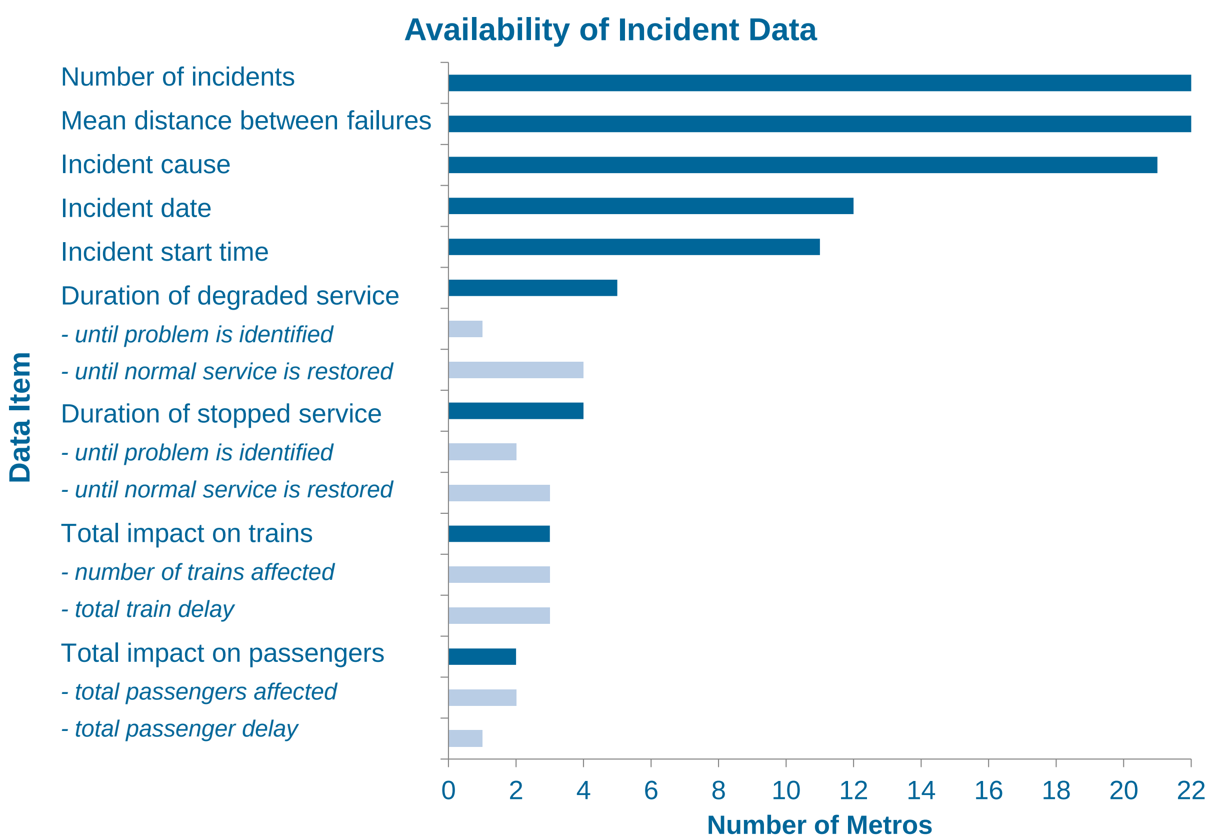

FIGURE 2 Overview of availability of incident data

One of the most common performance measures of service reliability used by CoMET and Nova metros is the mean distance between failures (MDBF). This indicator measures reliability in terms of incident frequency, regardless of how long a delay was incurred. The frequency of incidents lasting five minutes or more varies substantially among CoMET and Nova metros, as depicted in Figure 3. The figure shows how often incidents occur across metros, measured in terms of million car-km between incidents (MDBF). The x-axis shows the continent to which the metros belong to: Asia (As), Europe (Eu), Latin America (LA), and North America (NA). The y-axis is stretched to show the less reliable metros (to the right of the graph) in more detail.

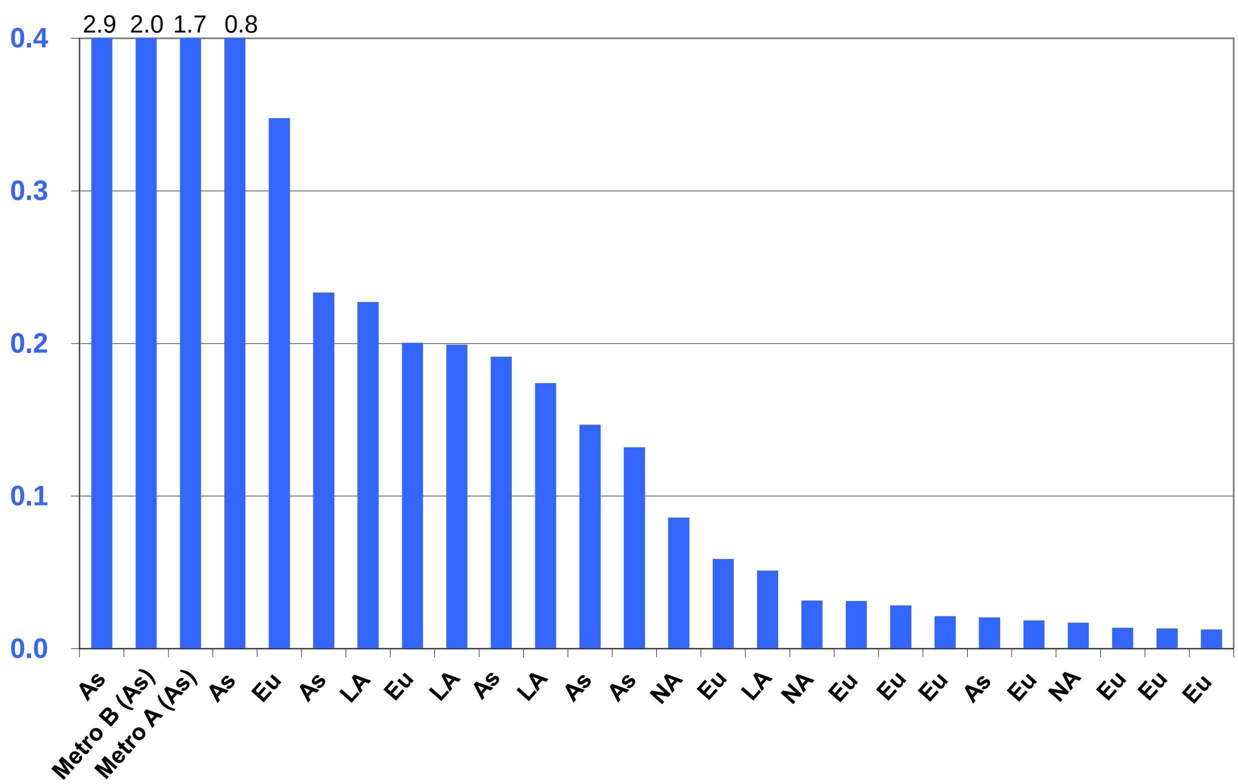

FIGURE 3 Million Car-km between 5-minute incidents (MDBF), 2010

Service reliability measures such as the mean distance between failures (MDBF) tells us how often incidents occur, but nothing about their impact. Total service reliability should also consider the length of delays and how many passengers are affected. In order to understand total service reliability we need measures that capture the impact of incidents on trains and customers, such as train hours delay and passenger hours delay. Moreover, we need to have these measures disaggregated by incident cause.

Of the two metros that were able to provide detailed incident data by cause, one was an extremely reliable metro with very few incidents per year (Metro B). As a result, too few data points were available to enable reliable analysis once the data were disaggregated by cause. The data from the other metro (Metro A) were therefore used for an extended analysis, which we present in the following section. The aim of this extended analysis was to test whether passenger-focused reliability indicators would lead to different conclusions from those suggested by incident frequency indicators, which in turn would influence how metros decide to best manage incidents and minimise passenger delay. Before we proceed to section 4.2, it is interesting to notice that Metro A and Metro B, the only two metros able to provide detailed data in the questionnaire, are two of the most reliable in CoMET and Nova (Figure 3).

4.2 Detailed analysis of train delay impact

An extended analysis of Metro A's incident data was performed to test the difference in conclusions reached when using passenger-focused indicators as opposed to frequency-based indicators and to then illustrate the consequences of making decisions based on incomplete data. The hypothesis was that the number of incidents is not a satisfactory proxy for passenger impacts, which can differ by type of incident (i.e. incident cause). This could be because some incident types take longer to resolve, some types have more knock-on effects, and some types are more common in crowded locations or during peak hours. As we will show, focusing solely on the number of incidents by cause can be misleading and cause operators to make sub-optimal decisions about the allocation of resources and investments for incident response and recovery.

Table 1 shows the results of this analysis when applied to six incident management performance indicators for Metro A. The indicators in the left-hand column are arranged from the least to the most passenger-focused. The final indicator, 'number of passengers affected by each incident type per year' is the result of multiplying the number of incidents of each type by the average number of passengers delayed by each incident type.

The table shows how the five incident types are ranked when classified by each indicator (1 = most severe). It is clear from the fact that four out of the five incident types are ranked most severe by at least one indicator, that the choice of indicator is crucial in determining which incident type appears to be most severe (and therefore influencing the management response). Rolling stock is the type of incident with the highest severity when we consider 'number of incidents' as the preferred indicator. Signalling is the type of incident with the highest severity when we consider 'number of trains affected', 'total train delay', and 'average number of passengers delayed', which is because signalling incidents are harder to resolve and therefore tend to last much longer than other incidents. Passenger-related incidents have the most severe impact on 'total passengers affected', which is believed to be largely due to the fact that these incidents tend to occur at busy times and locations. Other-equipment related incidents have the largest contribution to 'initial delay/resolution time', which is likely due at least partially to the inclusion of power-related incidents in this category. Finally, the table suggests that staff-related incidents have the least severe impacts according to all the indicators.

TABLE 1 Ranked severity of incident types depending on performance indicator use

Indicator	**Rolling Stock**	**Signalling**	**Other Equipment**	**Staff**	**Passengers**
Number of incidents (2009+2010 total)	**1**	3	4	5	2
Number of trains affected (per incident)	4	**1**	3	5	2

Initial delay/resolution time (per incident)	4	2	**1**	5	3
Total train delay time (per incident)	3	**1**	4	5	2
Average number of passengers delayed (per incident)	4	**1**	3	5	2
Total passengers affected by each type of incident (2009+2010 total)	2	3	4	5	**1**

Figure 4 shows a comparison of the most and least passenger-focused indicators. The primary (left) y-axis shows the total number of passengers affected in 2009 and 2010, and the secondary (right) y-axis shows the number of incidents in the same period. Incident types are shown on the x-axis. The figure shows that rolling stock incidents are nearly 5 times as numerous as staff and other equipment related incidents, 3 times more numerous than signalling incidents, and 1.3 times more numerous than passenger incidents.

As shown in Figure 2, 'number of incidents' is the primary performance indicator used by most respondents. If Metro A managed based on the number of incidents, they would see the biggest challenge as rolling stock. However, Figure 4 also shows the total number of passengers that were affected by each type of incident. The total number of passengers affected shows that if Metro A managed based on the total passengers affected they would see the biggest challenge as passenger incidents, not rolling stock. This supports the hypothesis that number of incidents is not an accurate proxy for effect on passengers. Therefore, use of a performance indicator that specifically addresses passengers is indeed a necessary pre-requisite for passenger-focused management of incidents.

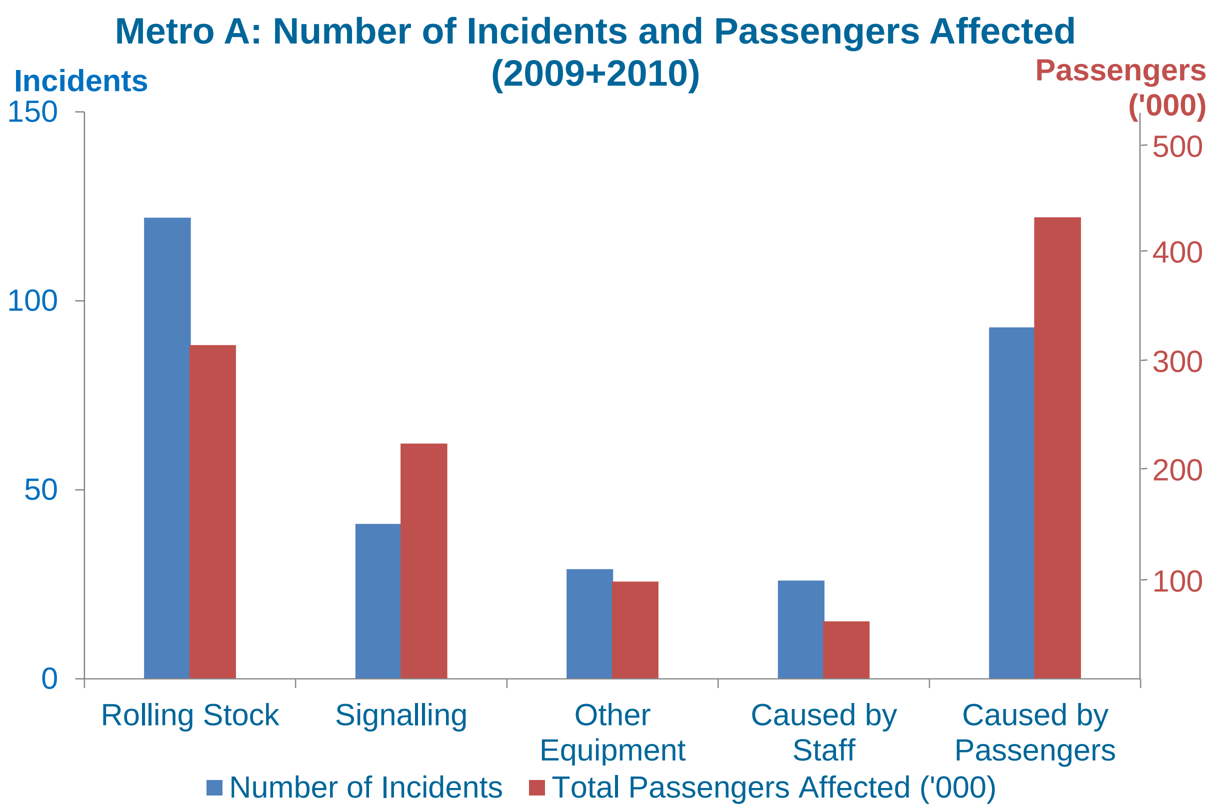

FIGURE 4 Number of incidents and passengers affected

Figures 5 to 7 illustrate other examples of how the variation in apparent significance of incident types varies according to the indicator used. The vertical y-axis in each figure shows: the average number of trains affected per incident (Figure 5); the average train delay per incident (Figure 6); and the average number of passengers delayed per incident (Figure 7). The horizontal x-axis shows incident types, for all three figures.

Signalling incidents on average affect more than twice as many trains as rolling stock incidents (Figure 5) and cause at least 50% more total train delay time per incident than passenger incidents (Figure 6). Signalling incidents also affect the most passengers per incident (Figure 7). Once again, these alternative performance indicators show a different result to the simplest indicator (number of incidents) and to the most passenger-focused indicator. This serves to re-emphasise that the performance indicator used affects metros' understanding of which type of incident is causing the greatest problem for their passengers. This in turn affects the decisions that are made in managing the response to incidents.

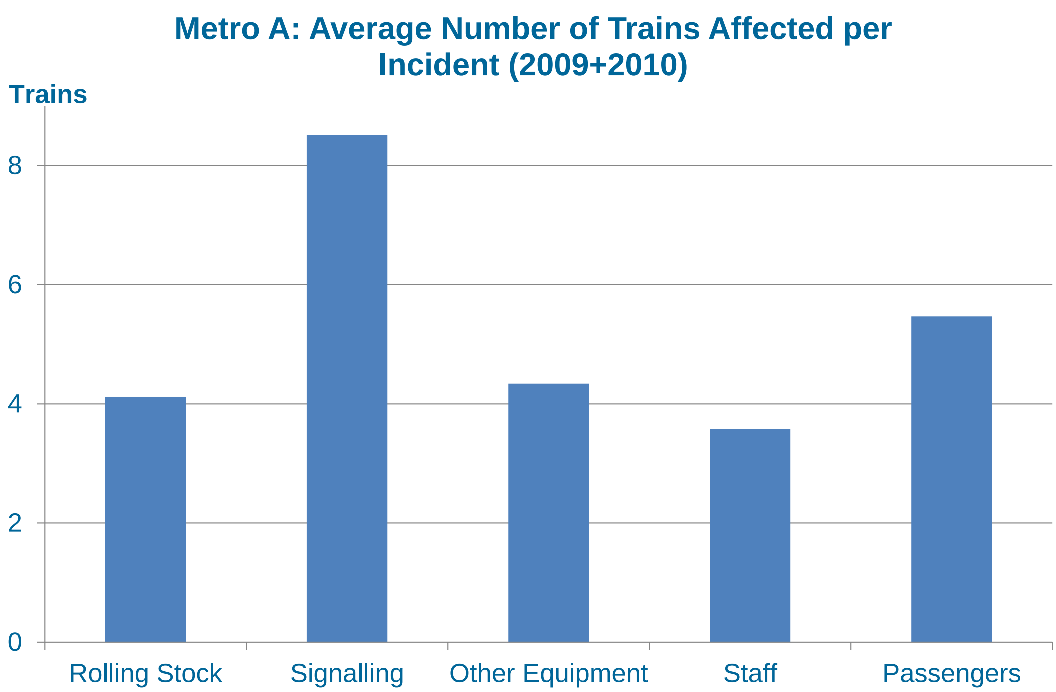

FIGURE 5 Average number of trains affected per incident

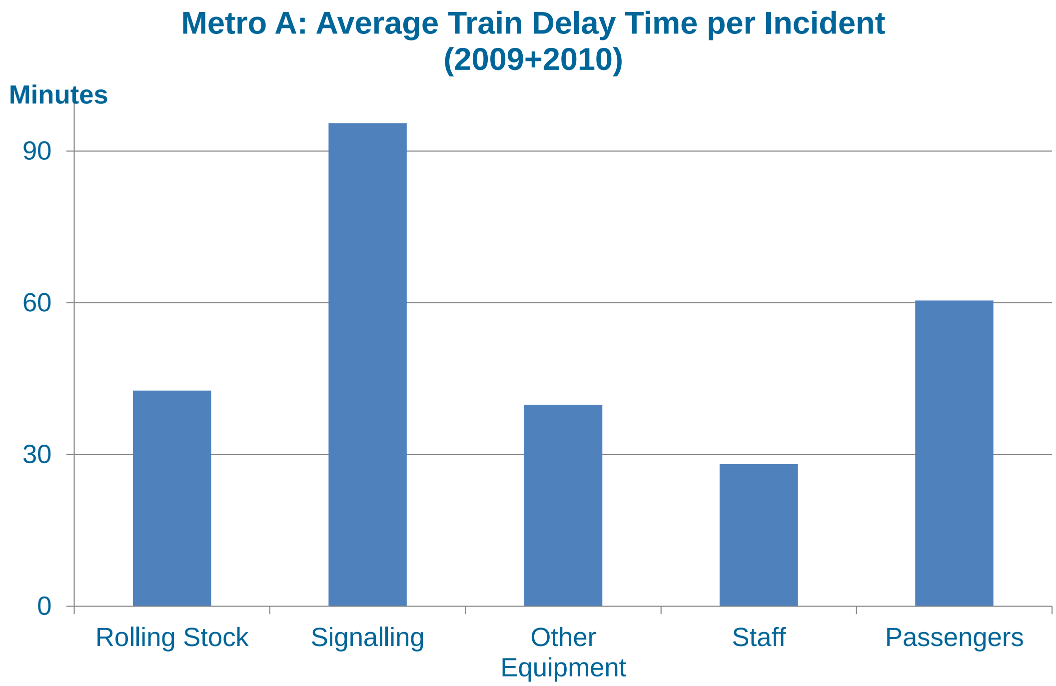

FIGURE 6 Average train delay time per incident

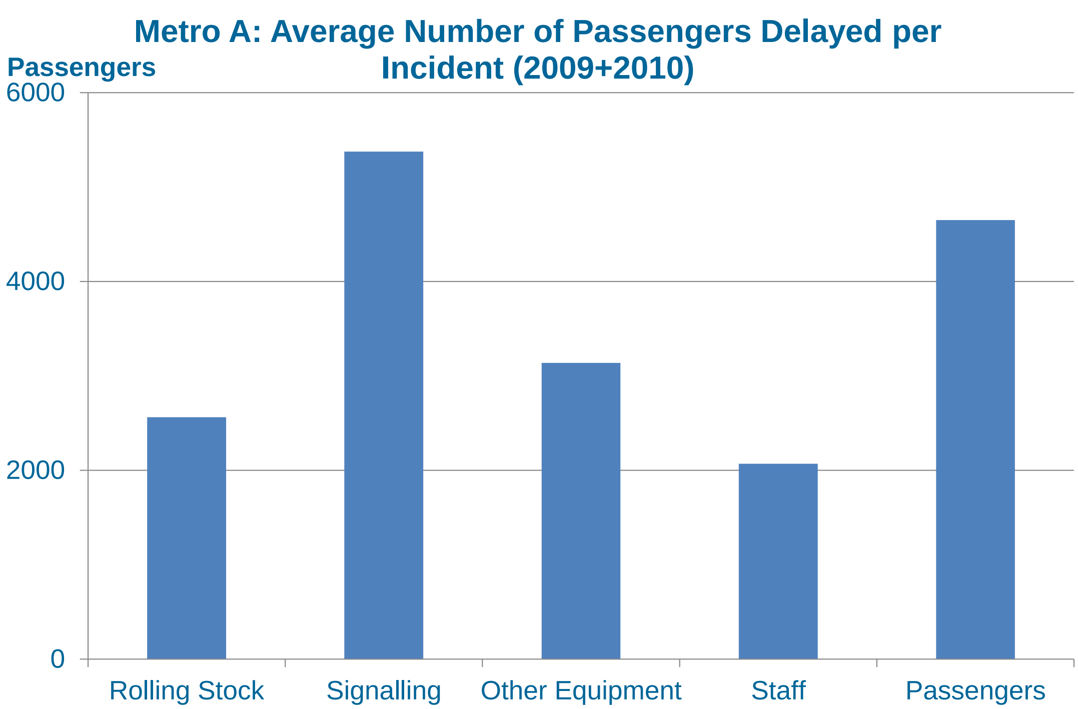

FIGURE 7 Average number of passengers delayed per incident

5. CONCLUSIONS AND RECOMMENDATIONS

Train delay incidents have major impacts on both operations and customers. Based on this research (see Figure 2) and the ongoing benchmarking experience from the CoMET and Nova metro groups, it is clear that most operators only measure and report the frequency of incidents. The common performance measures include incident rates and the mean distance between failures (MDBF).

However, these measures of incident frequency do not capture the full impacts of incidents, which depend on several factors including the cause of the incident, the duration, the location, the time of day, etc. Of the 22 metros who responded to the questionnaire, only two were able to provide sufficiently detailed data to enable an estimation of the total number of passengers affected to be calculated. This is partly due to the fact that many operators do not have the necessary data collection methodologies or systems in place. Even with good systems, calculating the full impacts is complex, especially in comparison to a simple count of incidents. The evidence shows that overcoming these measurement challenges is important, because the resulting performance measures provide a better understanding of how incidents impact customers. Therefore, operators need to collect and analyze a wide range of data to understand the impact of incidents and train delays on passengers' journeys.

The analysis conducted in this paper has showed that there are important differences between the frequency of incidents, duration of service delay, and degree of total disruption to trains and passengers by cause. For example, although rolling stock incidents occur more frequently in Metro A, passenger incidents actually affect more passengers. As a result, achieving a reduction in the frequency of rolling stock incidents may not have a major effect on the overall service reliability and customer satisfaction, if the incidents that have not occurred are those that have smaller impacts. Based on the data available for Metro A, it can be observed that a 10% reduction in the number of rolling stock incidents is associated with a reduction of less than 4% in the number of passengers affected per incident, while a 10% reduction in the number of signalling incidents is associated with a much larger reduction, about 24%, in the number of passengers affected per incident.

Based on this research, it is recommended that transit operators begin to measure the full impacts of incidents on trains and customers. At the aggregate level, measures of total hours of train delay and passenger delay should be the starting point. Passenger hours of delay, in particular, is the most passenger-focused measure of delay, but also the most difficult to measure. Analysis of this detailed data enables operators to get the best understanding of what is happening, facilitating continuous improvement. It is no coincidence that the only two metros able to provide this level of detailed data in the questionnaire are in fact two of the most reliable in CoMET and Nova.

The key benefit to be realized by measuring and analyzing the full impacts of incidents is an improvement in service quality, through a reduction in both the number of incidents and, more importantly, the overall duration and impact of incidents. Improved service reliability and fewer delays for passengers lead directly to higher performance and better customer satisfaction. The spin-off effects of these positive changes include improving the operator's reputation, increased support from the public and the government.

Operators may not be able to generate support for investments without being able to fully estimate the benefits, such as quantifying the operating cost savings as well as passenger time saved of potential investments. For example, what will the corresponding time/money savings of reducing average passenger delays by 5 minutes? Developing passenger-focused indicators enables transit managers to make more cost-effective decisions about when and where to direct resources.

As discussed in Section 2, an increasing number of countries are using cost-benefit analysis (CBA) to evaluate in monetary terms the benefits from reduced delays and improved travel time reliability. This can be done by using reliability multipliers and/or delay elasticities (see Section 2), which allow quantifying the economic impact of a given investment on patronage and fare revenue. By incorporating reliability into CBA, transit operators can make better informed decisions about alternative investments.

ACKNOWLEDGEMENTS

The authors wish to thank the CoMET and Nova benchmarking groups for participating in the data collection process, and Lucas Bueno dos Santos for his assistance.

REFERENCES

1. Rehnström, G., *Increasing the service quality of metropolitan railways*. 1991: Report 3. 49th International UITP Congress, Stockholm.
2. Lombart, A. and M. Favre, *Global quality of metros*. 1995: Report 2. 51st International UITP Congress, Paris.
3. Bates, J., P. Jones, and J. Polak, *The Importance of punctuality and reliability: a review of evidence and recommendations for future work*. 1995: Report to BR Business Consultancy, University of Westminster, Transport Studies Group.
4. Bates, J., P. Jones, J. Polak, and X.-L. Han, *The Investigation of punctuality and reliability: re-analysis of some existing datasets*. 1997, Report submitted to Rail Operational Research, University of Westminster, Transport Studies Group.
5. Van der Mede, P.H.J., H. Palm, and H. Flikkema, *Travel time variability as a new quality indicator*. 1996: Proceedings of Seminar E - part 2. 24th European Transport Forum, 2 - 6th Sept 1996.
6. Van Lint, J.W.C., H.J. Van Zuylen, and H. Tu, *Travel time unreliability on freeways: Why measures based on variance tell only half the story.* Transportation Research Part A: Policy and Practice, 2008. **42**(1): p. 258-277.
7. OECD, *Improving reliability on surface transport networks*. 2010: International Transport Forum, OECD, Paris.
8. Nicholson, A.J. and Z.P. Du, *Degradable Transportation Systems: An Integrated Equilibrium Model.* Transportation Research Part B: Methodological, 1997. **31(3)**: p. 209-223.
9. D'Este, G.M. and M.A.P. Taylor, *Network Vulnerability: An Issue for Regional, National and International Strategic Transport Networks*. 2001, Proceedings of 1st International

Symposium on Transportation Network Reliability (INSTR), July/August 2001, Kyoto University.

10. Lomax, T., D. Schrank, S. Turner, and R. Margiotta, *Selecting Travel Reliability Measures*. 2003, Texas Transportation Institute, Cambridge Systematics, Inc.
11. *Value of Travel Time Reliability and Cost-Benefit Analysis*: International Meeting Jointly organised by the Strategic Highway Research Program (SHRP2) of the Transportation Research Board and the Joint Transport Research Centre (OECD). 15-16 October 2009, Vancouver, Canada.
12. Bates, J., J. Polak, P. Jones, and A. Cook, *The valuation of reliability for personal travel.* Transportation Research Part E-Logistics and Transportation Review, 2001. **37(2-3)**: p. 191-229.
13. De Jong, G., E. Kroes, R. Plasmeijer, P. Sanders, and P. Warffemius. *The value of reliability*. in *Applied Methods Stream, European Transport Conference*. 2004. Strasbourg, 4-6 October 2004.
14. Preston, J., G. Wall, R. Batley, J.N. Ibanez, and J. Shires, *Impact of delays on passenger train services. Evidence from Great Britain.* Transportation Research Record: Journal of the Transportation Research Board, 2009. **No. 2117**: p. 14-23.
15. Vincent, M. and B.A. Hamilton, *Measurement valuation of public transport reliability*. 2008, Land Transport New Zealand Research Report 339.
16. Tseng, Y., P. Rietveld, and E. Verhoef, *A meta-analysis of valuation of travel time reliability*, in *Colloquium Vervoeresplanologisch Speurwerk*, *Part 3*, *CVS* 2008: Rotterdam.