Accident Analysis and Prevention 98 (2017) 232–240

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Modelling the impact of causal and non-causal factors on disruption duration for Toronto's subway system: An exploratory investigation using hazard modelling

Jacob Louie (M.Eng.)[a,*], Amer Shalaby (Ph.D., P.Eng.)[b], Khandker Nurul Habib (Ph.D., P.Eng)[b]

[a] *University of Toronto, Canada*
[b] *Department of Civil Engineering, University of Toronto, Canada*

ARTICLE INFO

Article history:
Received 8 February 2016
Received in revised form
21 September 2016
Accepted 5 October 2016
Available online 19 October 2016

Keywords:
Public transit
Subway
Service delay
Service disruption
Survival analysis

ABSTRACT

Most investigations of incident-related delay duration in the transportation context are restricted to highway traffic, with little attention given to delays due to transit service disruptions. Studies of transit-based delay duration are also considerably less comprehensive than their highway counterparts with respect to examining the effects of non-causal variables on the delay duration. However, delays due to incidents in public transit service can have serious consequences on the overall urban transportation system due to the pivotal and vital role of public transit. The ability to predict the durations of various types of transit system incidents is indispensable for better management and mitigation of service disruptions. This paper presents a detailed investigation on incident delay durations in Toronto's subway system over the year 2013, focusing on the effects of the incidents' location and time, the train-type involved, and the non-adherence to proper recovery procedures. Accelerated Failure Time (AFT) hazard models are estimated to investigate the relationship between these factors and the resulting delay duration. The empirical investigation reveals that incident types that impact both safety and operations simultaneously generally have longer expected delays than incident types that impact either safety or operations alone. Incidents at interchange stations are cleared faster than incidents at non-interchange stations. Incidents during peak periods have nearly the same delay durations as off-peak incidents. The estimated models are believed to be useful tools in predicting the relative magnitude of incident delay duration for better management of subway operations.

1. Introduction

In public transit systems, unforeseen and lengthy incidents and their consequential delays can adversely impact transit service, users' satisfaction with public transit and transit ridership in the long run. While complete prevention of incidents may never be possible, predictable delays and provision of upfront information to affected riders on the expected duration of service recovery can reduce the potential loss of ridership. From the standpoints of transit operations and planning, the ability to predict the possible durations of delay due to various causes can help the operating agencies decide whether or not an incident is disruptive enough to warrant a formal response (Schmöcker et al., 2005) and to help customers decide on alternative routes or services (Straphangers Campaign, 2014). Furthermore, although transit agencies have at their disposal measures to allow passengers to bypass service disruptions, such as bus bridging and diversionary tactics, such procedures often require an immense amount of scarce resources, and can only provide an inferior level of service. The Toronto Transit Commission (TTC) also notes that sourcing buses can cannibalise other service, creating problems elsewhere. Therefore, according to many transit agencies, minimizing the delay duration remains paramount over relying on such bypassing procedures (Pender et al., 2012).

Armed with detailed data on various types of incidents and observed delays in a recent year (2013), this paper focuses on developing statistical models that can quantify the impact of various attributes of a given incident on the delay duration in the TTC subway system, and to examine possible factors that aggravate a delay.

* Corresponding author.
E-mail addresses: jacob.louie@mail.utoronto.ca (J. Louie), amer@ecf.utoronto.ca (A. Shalaby), khandker.nurulhabib@utoronto.ca (K.N. Habib).

http://dx.doi.org/10.1016/j.aap.2016.10.008
0001-4575/

To accomplish this goal, a non-linear hazard model is fitted. The contributions of this study will be useful for the TTC and other similar transit agencies to determine the effect of peak periods, station types, and vehicle types. It should be emphasized that developing a model to generate a point-estimate prediction of a delay was not the goal of this study; however, the results of this study may contribute to the development of such a predictive model in the future.

The paper is organized as follows: Section 2 presents an overview of the performance of the Toronto subway system in recent years and some common causes for delays in subway operations. Section 3 presents a brief literature review of previous studies of incident duration in the transportation context. Section 4 describes the dataset as well as the formulation of the models and covariates. After a brief introduction to the hazard model in Section 5, Section 6 presents the capacity of the models in fitting observed data through diagnostic plots and goodness-of-fit statistics, as well as an interpretation of the parameter estimates of the final recommended model. The paper concludes with a summary of key findings, a discussion of the model's possible applications and limitations in subway system operations and management, as well as recommendations for policy and future research.

2. Trends in the Toronto subway performance

Toronto has a well-balanced transit system with a four-line subway network at its core, facilitating 41% of the 1.8 million daily transit trips, and carrying over 28,000 people per hour in the busiest direction (Yauch, 2014). As shown in Fig. 1, the TTC subway system extends to a total length of 68 km including 69 stations, five of which are interchange stations.

There were about 117 trainsets in service throughout the system, 47 of which were Toronto Rocket (TR) type trains (Toronto Transit Commission, 2013b). Launched in 2012, the TR trains are a new style of trains that are distinguishable by their open gangways between railcars, providing higher capacities than the traditional disjointed railcars. The TR trains were also the only trains equipped with a passenger intercom.

The Yonge-University-Spadina (YUS) line, the busiest subway line in Toronto shown in yellow in Fig. 1, was the only line that has consistently fallen short of its on-time performance target of 96% since at least 2013 (Toronto Transit Commission, 2015). This shortfall in performance could be due to a number of factors. Firstly, the YUS line has already been running above capacity during the morning rush hour (Yauch, 2014). Secondly, as the subway system ages and demand increases, the need to update the infrastructure becomes more urgent (Man et al., 2014). Work has been underway on the YUS line to upgrade the signalling system and other critical infrastructure. Thirdly, the new TR trains were the cause of many technical delays when they were first introduced (Toronto Transit Commission, 2013a).

In Toronto, the incident type that resulted in the most delay over the year 2013 was security, followed by door problems. The breakdown of delays by cause is given in Fig. 2a and b. Detailed definitions for each cause are given in Table 1. Other large cities experienced similar problems, although in different proportions. In New York City, the top delay causes include sickness, door holdings by passengers, speed-control and signal issues, as well as security. Some of these causes, such as police stoppages, have become more prevalent in light of terrorist attacks in recent history (Chan and McGinty, 2005).

As can be seen from the pie charts in Fig. 2a and b, lengthier delays do not necessarily correspond with more frequent delays. The variation of incident durations for each incident type is also very large, as shown in Fig. 2c.

3. Literature review

Although analysis of incident duration has been conducted extensively in the transportation context, much of the research is restricted to highway traffic only (Giuliano, 1989; Nam and Mannering, 2000; Valenti et al., 2000; Weng et al., 2014). Furthermore, past research found on duration analysis of public transit incidents were not at the level of detail that studies on highway incidents were able to achieve, as far as incorporating other non-causal variables is concerned (Weng et al., 2014). Several types of modelling approaches have been used to analyse the delay duration (Valenti et al., 2000), with hazard analysis (also known as Survival analysis) being the preferred approach in almost all of the transportation incident duration studies found (Chung, 2010; Giuliano, 1989; Nam and Mannering, 2000; Sharman et al., 2012; Weng et al., 2014). Three different highway-traffic studies had access to highly detailed datasets where the total duration time was decomposable into: time to detection, time to respond, time to travel to the incident, and time to clear the incident, with each time segment being analysed using separate hazard models (Chung, 2010; Nam and Mannering, 2000; Valenti et al., 2000). Conversely, the subway-based study did not break the total delay time into separate time components (Weng et al., 2014). Although different specifications of the hazard model were considered, the Log-Logistic specification was the most popular in the transportation context (Nam and Mannering, 2000; Sharman et al., 2012; Weng et al., 2014). Most studies found in both highway and transit fields also incorporated either unobserved heterogeneity (Nam and Mannering, 2000; Sharman et al., 2012) or mixed effects (Weng et al., 2014) to account for unobservable differences between observations.

Nearly all previous highway-based studies investigated the effect of non-causal variables, such as the time of day, the number of lanes affected around the scene, the size of the vehicle (Giuliano, 1989), weather characteristics, season, geographic location (Nam and Mannering, 2000), external response agencies involved (EMS, fire, or police), and the size of the response crew (Valenti et al., 2000). Conversely, the sole study found on subway system delays investigated only causal variables, such as power failure, vehicle failure, switch malfunction, and accidents, although incorporating non-causal variables, such as passenger volumes, was recommended for future investigations (Weng et al., 2014).

Temporal transferability of models between years, tested only for studies with multi-year data sets (Chung, 2010; Nam and Mannering, 2000; Weng et al., 2014), was found to be negatively affected by changes in governance or corporate structure (Weng et al., 2014) or changes in response procedures (Nam and Mannering, 2000). Nam and Mannering postulated that the incident management program of their subject city had not matured and was rapidly evolving at the time, causing the model results to be relatively unstable from year to year.

It is clear from this literature review that existing studies on public transit delays are nowhere near as comprehensive as studies on highway traffic delays, as far as incorporating non-causal variables and identifying the individual time segments are concerned. Incorporating non-causal variables, in addition to causal variables, is therefore a motivation for this study.

4. Data description, preprocessing, and methods

4.1. General data description

The TTC provided subway incident data only for the year 2013 for this investigation. For each record, the dataset includes the date, time, day of week, station, direction of travel, line, resulting delay in minutes, vehicle number, operator number, guard number, a

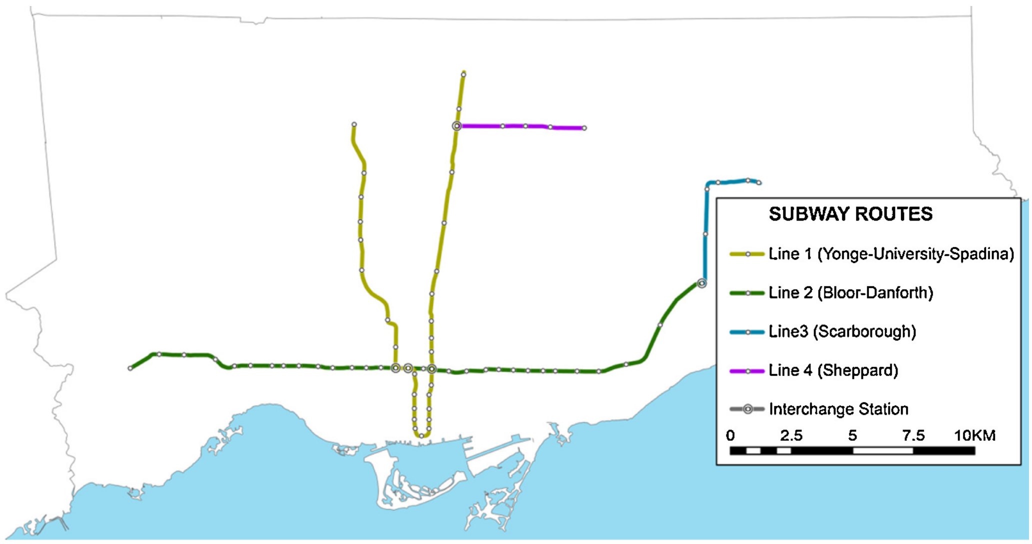

Fig. 1. Toronto Transit Commission Subway Map.

Source: (Binns, 2013).

Table 1
Definition of Variables Describing Time, Location, Vehicle Type Involved and Incident Type.

Variable Name	Definition
Peak_hour1	If the incident occurred during rush hour, from 6:00–9:00, or from 15:00–19:00, Monday–Friday
Interchange_station1	If the incident occurred at an interchange station
Intercom1	If the train type involved was a Toronto Rocket (TR) type
Proc_follow	If proper procedures were followed in response to the incident
Proc_follow1:Type	The type-specific interaction effect of adhering to proper procedures on the expected delay
TypeASSAULT	Incidents involving assault of any kind on passengers or personnel
TypeCOMMUNICATIONS	Problems with communication infrastructure with Transit Control
TypeDEBRIS___TRACK	Debris, objects, or other obstructions at track level
TypeDISORDERLY_PATRON	Disorderly patron
TypeDOOR	Mechanical door problems not caused by passengers
TypeDOOR_PASSENGER	Door problems caused by passengers
TypeDOOR_PERSONNEL_MISTAKE	Door opened off platform or in tunnel by personnel
TypeESCALATOR_ELEVATOR_STAIRS	Escalator/Elevator/Stair problem
TypeFIRE___TRACK	Fire at track level
TypeFIRE_IN_STATION	Fire elsewhere in station
TypeFIRE_ON_TRAIN	Fire on train
TypeHOLDUP_ALARM_ACTIVATED	Holdup alarm activated
TypeINJURY	Injury to personnel or customer
TypeNO_POWER	Loss of Power – includes traction and station power loss
TypeONBOARD_MECHANICAL_MAJOR	Major vehicle problems
TypeONBOARD_MECHANICAL_MINOR	Minor vehicle problems
TypeOPERATOR_NOT_AVAILABLE_NOT_IN_POSITION	Operator and/or guard not in position for duties
TypeOPERATOR_OVERSHOT_PLATFORM	Operator overshot platform
TypeOTHER	Other. This was used as the baseline category
TypePAA_ACTIVATED_BY_CUSTOMER	Passenger assistance alarm activated by customer or unknown person
TypePERSONNEL_ERROR	Mistake in operation by personnel or supervisor
TypeSECURITY_OTHER	Incidents involving security or police
TypeSIGNAL_SWITCH	Problems involving wayside signal or track switch problems
TypeSPEED_CONTROL	Speed control restricting operations – includes emergency brakes failing to release and gliding brakes
TypeSUICIDE_	Moving trains coming in contact with person
TypeTRACK_PROBLEM	Defective track structure
TypeTRACTION_POWER_PROBLEM	Traction power problem
TypeTRAIN_STOP_CONTACTED	Train tripped past a train stop
TypeTRAINING_DEPARTMENT	Incident caused by personnel in training
TypeUNAUTHORISED_TRACK_LEVEL	Unauthorised person at track level
TypeUNSANITARY_UNHEALTHY	Train taken out of service due to unsanitary conditions
TypeWEATHER	Incident caused by extreme weather
TypeWORK_REFUSAL	Delay caused by personnel refusing to work due to bad working environment
TypeWORKZONE_PROBLEMS	Delays caused by authorised track- level activity
TypeYARDHOUSE_PROBLEM	Incidents offline in the yardhouse

description of the incident, and a code representing the incident type.

There were around 12,600 recorded subway incidents in the year 2013, with the occurrence of incidents decreasing with increasing duration in a roughly exponential distribution. 36% of all incidents occurred during the morning and evening peak periods, defined as Monday–Friday between 6:00–9:00 and 15:00–19:00, respectively. 17% of incidents occurred at interchange stations. Finally, the Toronto Rocket (TR) trains accounted for 33% of all incidents.

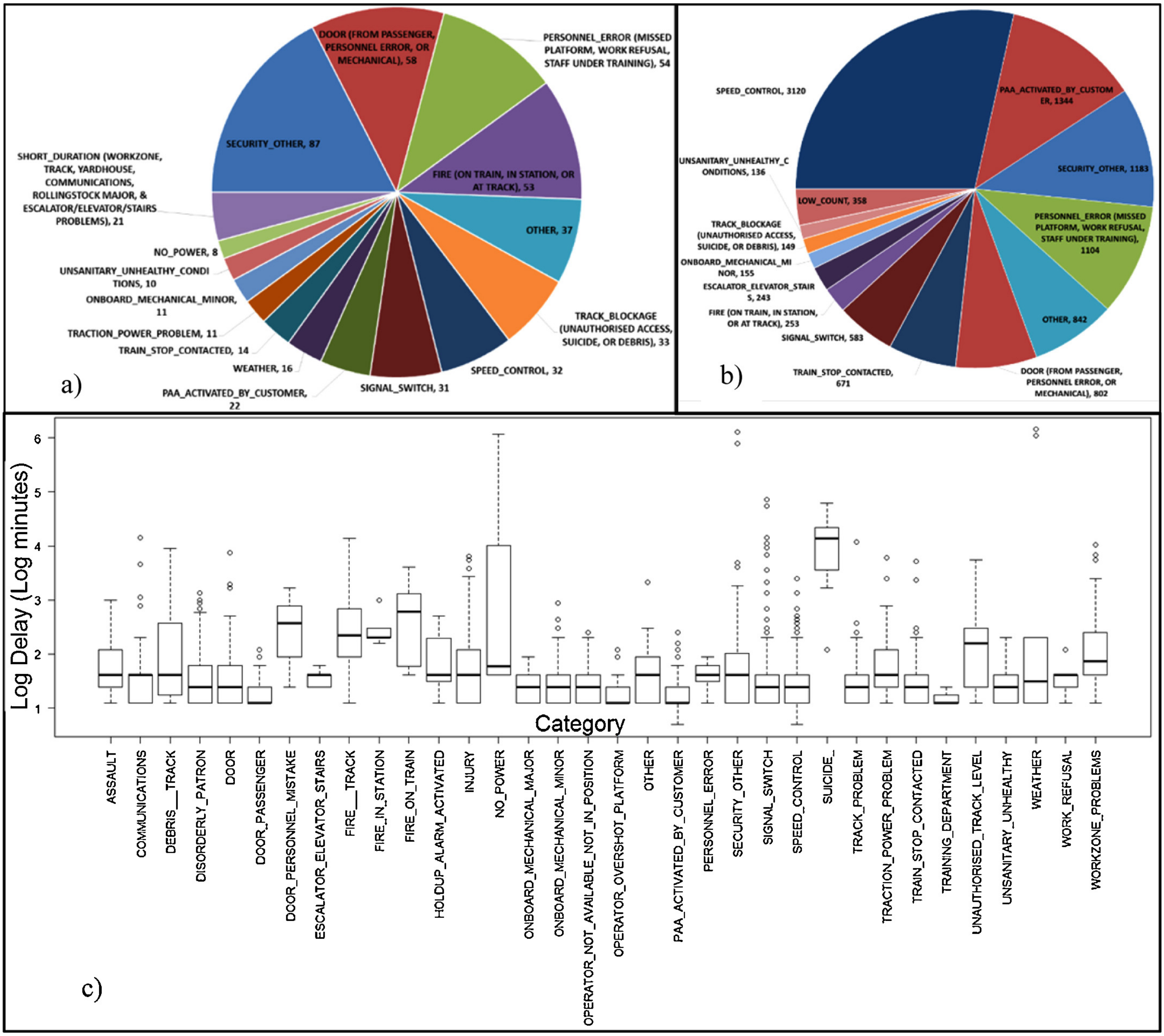

Fig. 2. Breakdown of incident types by: (a) Total duration (total number of hours of delay was 498) (b) Total count (Total number of incidents exceeded 12,000) (c) Distribution of duration delays for each incident type (plot excludes incidents of duration of less than 2 minutes).

55% of these incidents registered a delay of zero minutes, and the smallest non-zero delay was 2 min. It was unclear if delays less than 2 min were simply rounded down to zero minutes by the agency. It is also worth noting that other transit agencies, such as the MTA in New York City, the MTR in Hong Kong, and the MRT in Singapore, record only the incidents that exceed 8 min, 8 min, and 5 min, respectively (Straphangers Campaign, 2014; Weng et al., 2014). On the other hand, it can also be argued that short duration incidents of even less than 2 min can still be influential on the subway system's performance when headways are very short (Schmöcker et al., 2005). For the estimation of the hazard model, it was assumed that incidents with 0 min delay were actually left-censored observations below 2 min.

4.2. Variable definitions

Given the temporal and spatial distribution of incident occurrences discussed in the previous section, the effect of the following non-causal variables will be investigated: whether or not the train was a Toronto Rocket (TR) type train, whether or not the incident occurred at an interchange station, whether or not the incident occurred during rush hour, and whether or not proper procedures, if any, were followed. All non-causal variables were treated as binary dummy variables in this study, although different formulations of the variables were possible.

The descriptions field for each incident sometimes contained valuable information on the precise cause of the event and the reason for its prolongation, but it was not feasible to synthesize these details for all records. In this study, the descriptions field could be used to generate additional covariates only to the extent that there are recurring terms and phrases that relate groups of events with each other. For example, many incidents include the clause "Proper procedures followed". Therefore, another variable was created to indicate whether or not proper procedures, if any, were followed. Interaction terms between proper procedures and the causal variable were also estimated to examine the effectiveness of such procedures for different incident types.

The 100 codes which represent different incident causes were reducible to a minimum of 35 distinct categories without losing details on the nature of the incident. Previous researchers have devised their own formulations for the causal variable with fewer than 35 causes (Man et al., 2014; Weng et al., 2014); however, many of the incidents in the TTC's data set are not classifiable into these categories without losing important details on the severity of the incident.

The detailed definitions of all variables tested in this paper are given in Table 1. All variable names beginning with the prefix "Type" are causal variables – all other variables are non-causal variables.

5. Descriptions of model

5.1. Hazard analysis

Hazard models examine the conditional probability of an event occurring at any point in time, given that it has not yet occurred. In this study, the event is the termination of the incident delay. The hazard function is given in Eq. (1).

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \tag{1}$$

where h(t) is the conditional probability of incident termination at time t, and F(t) is the Cumulative Density Function of incident termination at time t. F(t) can be interpreted as the probability of the incident having been terminated before or at time t. S(t) is therefore the probability of the incident having survived to at least t. The hazard function itself can take on different parametric specifications, such as the Exponential, the Weibull, and the Log-Logistic specifications. An Exponential hazard model assumes a constant hazard over time, in which the conditional probability of incident termination does not change over time. Weibull hazard models allow for monotonically increasing or decreasing hazards. Log-Logistic models allow for a decreasing concave-up hazards, or hazards that increase to a maximum then decrease monotonically afterward (Allison, 2010; Machin et al., 2006).

Hazard analysis can also be used to examine the effect of covariates on the expected time to the termination of the incident itself, rather than the hazard of the incident termination, in the "Accelerated Failure Time" (AFT) formulation given in Eq. (2).

$$\log(T_i) = \beta_o + \sum_{j=1}^{Covariates} \beta_{i,j} * x_{i,j} + \beta_p * x_{i,p} + \sum_{k=1}^{IncidentTypes} \beta_{p,k} * x_{i,k} * x_{i,p} + \sigma * \varepsilon \tag{2}$$

For the i[th] incident, the symbols in Eq. (2) are defined as follows:

- β_o is the logarithm of the expected duration of the base incident type "Other" under base conditions (non-TR train at non-interchange stations during off-peak hours with no proper procedures followed).
- $\beta_{i,j}$ is the parameter for the j[th] covariate assuming proper procedures were not followed.
- $x_{i,j}$'s is 1 if the j[th] covariate for the i[th] incident is true, and 0 otherwise. The covariates in this summation include both causal and non-causal variables in Table 1, but does not include covariates involving proper procedures.
- β_p is the change in the expected log of the duration of the base incident type "other" if proper procedures were followed.
- $x_{i,p}$ is 1 if proper procedures were adhered to, and 0 otherwise.
- $\beta_{p,k}$ is the change in the parameter for the k[th] incident type over the base incident type if proper procedures were followed.
- ϵ is the random error term whose distribution depends on the specification, be it Exponential, Weibull, or Loglogistic.
- σ is the scale factor of the random error term that affects the specification of the AFT model.

Lengthier durations are predicted by positive parameter estimates in Eq. (2), whereas shorter durations are predicted by negative parameter estimates.

The aforementioned parametric specifications can also be expressed in AFT form. For an Exponential specification, ϵ has an extreme value distribution with the scale parameter σ being fixed to 1. For the Weibull specification, ϵ also has an extreme value distribution, but the scale parameter σ can take on any positive value. For the Log-Logistic specification, ϵ has a distribution as shown in Eq. (3).

$$PDF(\varepsilon_i) = \frac{e^{\varepsilon_i}}{(1 + e^{\varepsilon_i})^2} \tag{3}$$

Because the objective is to build a model that would examine the impact of various factors on the duration of the delay itself, rather than the probability of incident clearance over time, the AFT formulation is more relevant for this study's purposes than the hazard formulation. The AFT formulation was also chosen in other delay-duration studies in the transportation context (Nam and Mannering, 2000; Sharman et al., 2012; Weng et al., 2014). An advantage of AFT and Hazard modelling is that very few assumptions are imposed about the duration of very brief incidents, by allowing these observations to be left-censored at 2 min.

5.1.1. Heterogeneity

Although the effect of each covariate (i.e. the sign of the parameter estimates) on the resulting duration is not expected to vary between observations, the magnitudes of their effects may vary slightly between observations. This variability is considered through unobserved heterogeneity and can possibly stem from differences between each particular operator, guard, or railcar involved.

In the software R, a maximum of one heterogeneity term to replace one variable is allowed. The operator ID number was chosen to be replaced by the heterogeneity term, as it resulted in the lowest log-likelihood and AIC fit statistics. Heterogeneity is incorporated into the model by Eq. (4) (Nam and Mannering, 2000; Sharman et al., 2012).

$$S_\theta(t_i) = \int_0^\infty S(t_i)^\nu * g(\nu_i) * d\nu \tag{4}$$

where ν is the effect of the random variable —in this case, the operator ID number, and g(ν_i) is the assumed distribution of the random effect ν. If ν is assumed to follow the often-used gamma distribution (Nam and Mannering, 2000; Sharman et al., 2012), the survival function reduces to Eq. (5)

$$S_\theta(t) = \left[1 - \theta \ln\left\{S(\mathrm{t})\right\}\right]^{-\frac{1}{\theta}} \tag{5}$$

where θ is the variance of the random effect v. Heterogeneity plays a significant role in the model if θ is found to be significantly greater than 0 using the Wald and Likelihood Ratio tests, as explained in (Therneau et al., 2003).

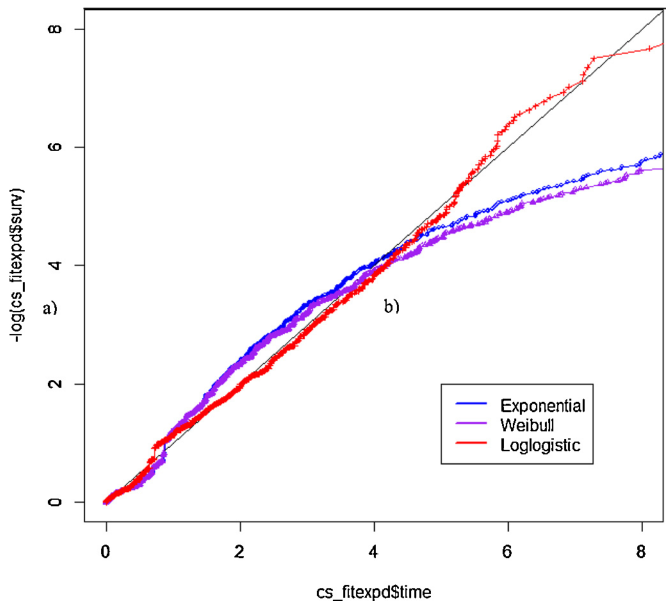

Fig. 3. Cox-Snell residual diagnostic plot for the Exponential, Weibull, and Log-Logistic AFT specification.

6. Discussion of results

6.1. Model appraisal

The adequacy of the proposed AFT model specification was checked by examining the graphical plot of the Cox-Snell residuals, shown in Fig. 3. For a specification to be appropriate, the resulting plot should closely align with a straight line having a slope of 1 and an intercept of 0 (Machin et al., 2006). The Exponential, Weibull, and Log-Logistic specifications were tested.

The Log-Logistic specification remains the best out of the three parametric specifications, as the slope of the residuals plots is closest to 1. The competing specifications can also be compared based on other fit statistics, such as the AIC. The Log-Logistic specification also had the lowest AIC value (at 33965.82), compared to the Exponential (at 37059.54), and the Weibull (at 36232.78).

For the final goodness-of-fit measure, the adjusted approximate McFadden's ρ^2 in Eq. (6) was used because the number of observations used was very large, which would have dampened the effect of including too many covariates (Sharman et al., 2012).

$$\rho^2{}_{adjusted} = 1 - \frac{\left(LL_{full}\right) - r}{LL_{null}} = 1 - \frac{-17613.6 - 53\ covariates}{-21608.3} = 0.182 \tag{6}$$

This $\rho^2{}_{\text{adjusted}}$ is several orders of magnitude larger than the $\rho^2{}_{\text{adjusted}}$ value calculated by Sharman, et al. (Comparison of Parametric and Nonparametric Hazard models for Stop Durations on Urban Tours with Commercial Vehicles, 2012) in their study, who also recommended the Log-Logistic AFT as the final model.

The Log-Logistic AFT is also consistent with the findings of other AFT studies in transportation (Sharman et al., 2012; Valenti et al., 2000; Weng et al., 2014). The fact that the scale parameter is less than 1, at 0.384, suggests that the hazard increases to a maximum, and then decreases afterward (Allison, 2010; Machin et al., 2006). An increasing hazard means that an event is increasingly likely to be cleared as time progresses. A decreasing hazard means that an incident is decreasingly likely to come to an end as time progresses, and this is analogous to incidents with extremely long durations. These conclusions on the time-dependency of the hazard point to a very wide spread of durations within each incident type, which is supported by the boxplots of Fig. 2c. These findings, along with the diagnostic plots throughout Section 6.1, support the AFT Log-Logistic specification as the final model to be used. The parameter estimates from the Log-Logistic AFT model are shown in Table 3. All variables are retained in the final model, whether they are statistically significant or not.

6.2. Effect of non-causal variables

From Table 3, incidents are cleared 5% faster on average at interchange stations than at other stations. It is likely that the TTC assigned higher priority at clearing incidents at interchange stations than at non-interchange stations, given that incidents at interchange stations potentially impact a larger portion of the network, and that interchange stations experience higher passenger volumes, as well as higher incident rates. Furthermore, most interchange stations are centrally located, making it easier for response crews to focus their attention on fewer locations across the network.

The coefficient for the effect of peak-hours is positive, but small in magnitude and statistically insignificant. Although it may have been expected that passenger and traffic congestion during peak periods hindered response crews from reaching the scene, the effect was hardly severe.

The coefficient for the effect of the involvement of a TR train is significantly positive, associated with delays that were 11% longer than those of non-TR trains. It was initially expected that the TR trains should have been associated with shorter delays, since they were the only trainsets equipped with a passenger intercom. This was because the duration of an incident can be highly dependent on the manner and ease of detection and identification, to which an intercom system provides another means (Roh [no date]). This unexpected result suggests that the presence of an intercom was not enough to overcome the other problems that the TR trains experienced in 2013. Furthermore, it is possible that passengers were not yet accustomed to using the newly available intercom effectively. In future years, however, the effect of TR trains may no longer increase the expected duration to the same degree as in 2013, as technical bugs from new technology become resolved over time. Therefore, the positive parameter estimate is likely more associated with the unfamiliarity of the glitches that accompany the introduction of any new technology, rather than the TR trains themselves.

Of incident types that have proper procedures in place, adherence to proper procedures were generally associated with shorter expected delays, for events in which such procedures were applicable. This can mean that either the adherence of proper procedures reduced the delay, or that incidents in which proper procedures were applicable tended to be brief. The exceptions to this are incidents involving weather, trains overshooting the platform, and power loss. These events account for only 200 of over 12,600 observations. It was possible that the duration of these incident types depended very little on whether or not proper procedures were applied, but on the severity of the incident itself of each incident type. This is supported by the fact that none of the interaction terms that were positive were statistically significant, suggesting that adherence to proper procedures mattered very little in the resulting total delay duration for these incident types. It was not possible in the dataset to distinguish the part of the duration due to the application of proper procedures themselves.

Of incident types for which proper procedures reduced the delay, the reduction in delay was more limited if the cause involved problems with passengers, be it due to injuries, alarm activations, or security. A possible explanation is that the transit agency has less control over external causes like passengers and more control over internal causes, such as its equipment; consequently, externally-induced delays may have been harder to diagnose than internally-induced ones.

Table 2
Testing the Value of Non-Causal Variables in the Final Model.

Variable	LL_o	LL_a	Dof	LR	X^2	Add Value?
Interchange_station1	−17971.2	−17967.4	1	7.6	3.841	Yes
Peak_hour1	−17967.4	−17966.2	1	2.4	3.841	No
Intercom1	−17967.4	−17949.1	1	36.6	3.841	Yes
Proc_follow1	−17949.1	−17546.4	1	805.4	3.841	Yes
Proc_follow1:Type	−17546.4	−17428	16	236.8	26.296	Yes

The value of including the statistically significant non-causal variables in the final model is tested sequentially using the Likelihood Ratio test in Eq. (7)

$$LR = -2\left(LL_o - LL_a\right) \sim \left(\chi^2_{dof=1,\alpha=0.05}, \chi^2_{dof=16,\alpha=0.05}\right) = (3.841, 26.296) \quad (7)$$

where LL_o is the log likelihood of the model without these variables, LL_a is the log likelihood of the model with the variables, and dof is the difference in the number of terms with and without the variables being tested. The results of the calculations are shown in Table 2.

All of the statistically significant non-causal variables, except for the effect of peak-hours, add value to the final model. Even though the TTC can adopt policies that are specific to the different causes and types of incidents, this conclusion supports the adoption of more overreaching policies that need not be cause-specific as well.

Whilst the model should not be used to generate actual point estimates of delays as explained in the conclusion, an expected delay was calculated only to help to quantify the net effect of adhering to proper procedures on the incident duration for each incident type. The calculated delays should not be interpreted as predicted delays for each incident type. For each incident type, the calculated base delays (off-peak, involving a non-TR train at non-interchange stations) without proper procedures and with proper procedures were calculated by Eq. (2) and assuming ε to be 0, with the results shown in columns 3 and 4 of Table 3.

As for the effect of who the operator involved was, which had been represented by the gamma heterogeneity term as discussed in Section 5.1.1, the variance of this random effect was significantly greater than 0, as seen at the bottom of Table 3. This suggests all individual staff members are not equally competent at responding to incidents. The fact that the model achieved the best fit when the operator ID variable was replaced by the heterogeneity term suggests that the operator is more influential on the delay duration than the guard or specific railcar involved. One possible reason may have been because operators could spot directly from the front of the train the cause of many incidents that impact operations, such as wayside signalling or track-borne problems. This suggests that the detectability of an incident plays an important part in the delay duration and should be further investigated in future studies.

6.3. *Effect of the causal variable*

As can be seen from Table 3, incidents that impact simultaneously train operations and safety generally have the longest duration. These incidents include workzone problems, suicides, fires on trains, fires on tracks, doors opening off the platform, and unauthorised access to track level. This result may reflect the TTC's "safety first" attitude that requires them to do a more thorough job of responding to life-threatening incidents. These results are also consistent with the findings of Weng et al. (Development of a Subway Operation Incident Delay Model Using Accelerated Failure Time Approaches, 2014), who explained that crews responding to incidents involving potential casualties have to prioritise rescue operations of human lives over the salvaging of property and equipment. Notable is the wide spread of expected delay durations amongst incident types of this group – incident types that involved actual casualties or injuries from moving trains coming in contact with people at track level last almost an order of magnitude longer than incidents involving unauthorised track-level access without fatalities or injuries.

Conversely, incidents that involve operations or safety alone have shorter delays. Safety incidents tend to be slightly shorter than operational incidents, but the difference is hardly obvious. Although a high importance may have been placed on health-and-safety incidents, medical incidents can often be treated off site in an ambulance or at the hospital. On the other hand, incidents involving police or fire services usually must be investigated and remedied on site. Furthermore, harmed customers often refuse aid if the injury is minor. Examining the reasons behind lengthier medical delays suggests that medical delays are aggravated if the affected person cannot be removed from the train, counteracting the factors that tend to shorten safety-only incidents. Furthermore, it was possible that safety-only incidents, usually caused by external influences such as customers, were more difficult to diagnose and troubleshoot than an operational-only incident, the latter usually being caused by an internal influence that largely stems from the transit agency itself.

Although power loss incidents can heavily affect operations, train operators are often instructed to bypass the affected station rather than stop if the power loss is not at track level. Power loss events were confounded with track power-off situations in the dataset, and the only detail that would distinguish these two types of events is embedded in the event descriptions. It was therefore not feasible to automatically distinguish between these two types of power-off incidents.

7. Conclusion

This paper presents an empirical model that can predict the effect of the incident characteristics on the duration of delays to subway operations in Toronto. An Accelerated Failure Time (AFT) model was estimated, and had satisfactory fit and diagnostic statistics. It also did not require any imposing assumptions about the duration of incidents less than 2 min. The parameter estimates were assessed and interpreted, and the results were generally explainable, and were found to be consistent with prior expectations and previous studies (Weng et al., 2014). Although incident duration in subway systems has been studied in the past using incident type as the explanatory variable, this study was able to determine that the effects of other non-causal variables such as station type, train type, and adherence to proper procedures on incident duration were statistically significant, and add value to the final prediction of their effect on delay duration. Given the importance of these non-causal variables, the TTC should endeavour to identify additional non-causal details, such as passenger volumes, in their datasets to allow for further investigation. The empirical model developed and presented in this paper can serve as an analytical tool for better management of the negative impacts of incident delays. In light of the study's findings, it is recommended that the transit agency:

Table 3
Parameter Estimates and Calculated Durations of the Final Model.

Covariate Name	Parameter Estimate	Base Delay (minutes)	Delay With PP	Safety (S), Operations (O), or both (SO)	Pct% change With PP
(Intercept)	1.08***	2.96			
TypeSUICIDE_	2.9***	53.5	NPP	SO	NPP
TypeFIRE_ON_TRAIN	1.41***	12.07	NPP	SO	NPP
TypeDOOR_PERSONNEL_MISTAKE	1.2***	9.79	6.33	SO	−35.4%
TypeFIRE___TRACK	1.19***	9.72	NPP	SO	NPP
TypeWORKZONE_PROBLEMS	0.78***	6.46	NPP	SO	NPP
TypeUNAUTHORISED_TRACK_LEVEL	0.63***	5.57	NPP	SO	NPP
TypePERSONNEL_ERROR	0.6***	5.41	0.79	O	−85.5%
TypeTRACTION_POWER_PROBLEM	0.6***	5.36	NPP	O	NPP
TypeOPERATOR_NOT_AVAILABLE_NOT_IN_POSITION	0.59***	5.34	2.16	O	−59.7%
TypeYARDHOUSE_PROBLEM	0.44***	4.57	NPP	O	NPP
TypeCOMMUNICATIONS	0.38***	4.33	NPP	O	NPP
TypeONBOARD_MECHANICAL_MINOR	0.27***	3.89	NPP	O	NPP
TypeDOOR	0.27***	3.88	NPP	O	NPP
TypeONBOARD_MECHANICAL_MAJOR	0.27*	3.87	NPP	O	NPP
TypeUNSANITARY_UNHEALTHY	0.25***	3.81	NPP	S	NPP
TypeDEBRIS___TRACK	0.23	3.74	1.21	O	−67.6%
TypeFIRE_IN_STATION	0.22	3.69	NPP	S	NPP
TypeSECURITY_OTHER	0.22***	3.68	2.96	S	−19.6%
TypeASSAULT	0.16***	3.48	NPP	S	NPP
TypeDISORDERLY_PATRON	0.09**	3.24	NPP	S	NPP
TypeWEATHER	0.08	3.22	25.32	O	687.1%
TypeDOOR_PASSENGER	0.08	3.2	NPP	O	NPP
TypeINJURY	0.02	3.01	2.46	S	−18.2%
TypeSPEED_CONTROL	−0.04	2.85	0.64	O	−77.7%
TypeWORK_REFUSAL	−0.09	2.71	NPP	O	NPP
TypeTRACK_PROBLEM	−0.1	2.68	2.67	O	−0.2%
TypeSIGNAL_SWITCH	−0.15***	2.55	1.61	O	−36.9%
TypeNO_POWER	−0.16	2.52	7.29	O	188.7%
TypeTRAINING_DEPARTMENT	−0.34	2.11	1.16	O	−44.7%
TypeTRAIN_STOP_CONTACTED	−0.39**	2.01	1.35	O	−33%
TypeOPERATOR_OVERSHOT_PLATFORM	−0.63**	1.57	1.74	O	11.1%
TypePAA_ACTIVATED_BY_CUSTOMER	−0.79***	1.34	1.16	SO	−13.5%
TypeHOLDUP_ALARM_ACTIVATED	−1.16***	0.93	NPP	S	NPP
TypeESCALATOR_ELEVATOR_STAIRS	−1.16***	0.92	NPP	S	NPP
Interchange_station1	−0.06***	NA	NA	NA	NA
Intercom1	0.11***	NA	NA	NA	NA
Proc_follow1	−0.35***	NA	NA	NA	NA
Peak_hour1	0.01	NA	NA	NA	NA
TypeDEBRIS___TRACK:Proc_follow1	−0.77	NA	NA	NA	NA
TypeDOOR_PERSONNEL_MISTAKE:Proc_follow1	−0.08	NA	NA	NA	NA
TypeINJURY:Proc_follow1	0.15	NA	NA	NA	NA
TypeNO_POWER:Proc_follow1	1.41**	NA	NA	NA	NA
TypeOPERATOR_NOT_AVAILABLE_NOT_IN_POSITION:Proc_follow1	−0.56	NA	NA	NA	NA
TypeOPERATOR_OVERSHOT_PLATFORM:Proc_follow1	0.46	NA	NA	NA	NA
TypePAA_ACTIVATED_BY_CUSTOMER:Proc_follow1	0.21	NA	NA	NA	NA
TypePERSONNEL_ERROR:Proc_follow1	−1.58***	NA	NA	NA	NA
TypeSECURITY_OTHER:Proc_follow1	0.13	NA	NA	NA	NA
TypeSIGNAL_SWITCH:Proc_follow1	−0.11	NA	NA	NA	NA
TypeSPEED_CONTROL:Proc_follow1	−1.15***	NA	NA	NA	NA
TypeTRACK_PROBLEM:Proc_follow1	0.35	NA	NA	NA	NA
TypeTRAIN_STOP_CONTACTED:Proc_follow1	−0.05	NA	NA	NA	NA
TypeTRAINING_DEPARTMENT:Proc_follow1	−0.24	NA	NA	NA	NA
TypeWEATHER:Proc_follow1	2.42	NA	NA	NA	NA
Scale	0.384	NA	NA	NA	NA
Variance of Random Effect	0.468	NA	NA	NA	NA

Significance Confidence Levels:*** 99%, ** 95%, * 90%, PP: Proper Procedures, NPP: Proper procedures were never applied for this incident type, Pct change With PP: Percentage change in delay duration if proper procedures are applied with negative being a decrease, NA: Not Applicable.

- Prioritise the prevention of incidents that simultaneously impact safety and operations,
- Focus on finding more ways of removing the source of the incident away from the line or train,
- Increase attention to clearing incidents at non-interchange stations, whilst maintaining the quick response performance at busier interchange stations,
- Establish official procedures for incident types that do not yet have such procedures,
- Increase the adherence to proper procedures through better training of personnel, and
- Advertise heavily the appropriate use of any new safety feature, such as a passenger intercom, which could help in identifying the cause of the incident.

Because this model could not be tested for temporal transferability due to the lack of data from other years, it should be used only to predict the effects of characteristics that are not likely to change of an incident on the delay duration, and should not be used to generate point-estimate predictions for a future year. One reason is that the effect of non-causal variables can change with time. For example, as the TTC becomes more familiar with resolving technical bugs of the new TR trains over time, it is unlikely that TR trains will continue to contribute to the same delays to the same degree as during their first two years of operation. Another reason is that changes in policy decisions and the response environment are likely to change the effect of variables in future years. For example, if the TTC decides to reallocate its resources to clearing incidents at non-interchange stations, this can result in longer delays at interchange stations during another year. Thirdly, aside from the issue of temporal transferability, the presence of heterogeneity also precludes point-estimation using the model. As was postulated that the heterogeneity stemming from who the operator was is related to the competency of the operator at responding to different incidents, perhaps the TTC can also include the competency level of each guard and operator involved for each incident in a quantifiable form, such as the number of years of experience. Notwithstanding the difficulty in generating actual duration predictions, it is expected that the inferences on the relative durations between each type of incidents predicted by the model will remain valid over time and across personnel – that incidents simultaneously involving both safety and operations will always have the longest durations, as response crews must dedicate time to prioritise the safety of lives in addition to salvaging equipment (Weng et al., 2014).

Although generating accurate predictions for delay duration remains a pressing need for transit agencies, another obstacle to the development of a predictive model was data imbalance. The data set was dominated by brief incidents, with very few lengthy incidents. Additionally, some incident types were far more prevalent than others. While this can present problems in training or fitting the model to predict durations of lengthier incidents or seldomly-occuring incident types, some data-balancing strategies exist, including sampling from the dataset for more equal proportions of incident durations and types for use in model training (Brownlee, 2015). One issue is the exclusion of a portion of data from training to retain for validation; however, the feasibility of constructing a predictive model can be investigated in future studies with more data, possibly from other years. Nevertheless, the results of this exploratory study may help to advance the goal of developing such a predictive model in the future.

Despite this study's accomplishments, the level of detail still falls short of what highway-based studies were able to achieve, in that the different phases of the total delay duration were not distinguishable in the TTC's data set. Achieving this level of detail would be possible if the TTC spells out precisely the definition of delay onset and clearance and uses automated incident detection systems (e.g. cameras) to distinguish the precise time of the start and end of the "detection" phase of the duration (Nam and Mannering, 2000). Once the TTC installs the necessary detection infrastructure, this can be the subject of future investigations.

Acknowledgements

The authors would like to thank all members of the ITS lab at the University of Toronto and colleagues at ARUP who have exchanged their ideas.

References

Allison, P.D., 2010. Survival Analysis Using SAS: A Practical Guide. SAS Institute.
Binns, P., 2013. TTC Subway Map to Scale Version Concept [Online] Available at: http://cranialmatter.com/dt_portfolio/ttc-subway-map-to-scale-version-concept (accessed 28.07.15.).
Brownlee, J., 2015. 8 Tactics to Combat Imbalanced Classes in Your Machine Learning Dataset [Online] Available at: http://machinelearningmastery.com/tactics-to-combat-imbalanced-classes-in-your-machine-learning-dataset/ (accessed 18.07.16.).
Chan, S., McGinty, J., 2005. Think the Subway's Running Later? You're Right [Online] Available at: http://www.nytimes.com/2005/03/26/nyregion/26subway.html?pagewanted=all (accessed 2014).
Chung, Y., 2010. Development of an accident duration prediction model on the korean freeway systems. Accid. Anal. Prev. 42 (1), 282–289.
Giuliano, G., 1989. Incident characteristics, frequency, and duration on a high volume urban freeway. Transp. Res. Part A: Gen. 23 (5), 387–396.
Machin, D., Cheung, Y.B., Parmar, M., 2006. Survival Analysis: A Practical Approach. Wiley & Sons, Chichester, UK.
Man, J., Misra, A., Shalaby, A., 2014. A Survey-Based Approach to Understanding the Resilience Implications of Day-To-Day Disruptions on Transit Operations (Unpublished Report). University of Toronto, Toronto, Ontario, Canada.
Nam, D., Mannering, F., 2000. An exploratory hazard-based analysis of highway incident duration. Transp. Res. Part A: Policy Pract. 34 (2), 85–102.
Pender, B., Delbosc, A. & Shiwakoti, N., 2012. Planning for the Unplanned: An International Review of Current Approaches to Service Disruption Management of Railways.
Roh, S.K., 2014. A study on the emergency response manual for urban transit fires. Proceedings of the 7th Asia-Oceania Symposium on Fire Science and Technology, n.d. [Online] Available at: http://www.iafss.org/publications/aofst/7/17/view (accessed 2014).
Schmöcker, J.D., Cooper, S., Adeney, W., 2005. Metro service delay recovery: comparison of strategies and constraints across systems. Transp. Rese. Rec.: J. Transp. Res. Board, 30–37.
Sharman, B., Roorda, M., Habib, K., 2012. Comparison of parametric and nonparametric hazard models for stop durations on urban tours with commercial vehicles. Transp. Res. Rec.: J. Transp. Res. Board 2269 (Issue), 117–126.
Straphangers Campaign, 2014. Methodology: Straphangers Campaign Analysis of MTA Alerts of Subway Incidents/Delays in 2011 and 2013 [Online] Available at http://www.straphangers.org/alerts/14/Methodology.pdf.
Therneau, T.M., Grambsch, P.M., Pankratz, V.S., 2003. Penalized survival models and frailty. Am. Stat. Assoc.
Toronto Transit Commission, 2013a. Chief Executive Officer's Report – February 2013 Update, Toronto.
Toronto Transit Commission, 2013b, [Online] Available at: https://ttc.ca/About_the_TTC/Operating_Statistics/2013.jsp (accessed 29.07.15.).
Toronto Transit Commission, 2015. Chief Executive Officer's Report – January 2015 Update [Online] Available at: https://www.ttc.ca/About_the_TTC/Commission_reports_and_information/Commission_meetings/2012/January_2/Report/CHIEF_EXECUTIVE_OFFICER_S_REPORT_JANUARY_2015_UPDATE.pdp (accessed 2015).
Valenti, G., Lelli, M., Cucina, D., 2000. A comparative study of models for the incident duration prediction. Eur. Transp. Res. Rev. 34 (2), 85–102.
Weng, J., Zheng, Y., Yan, X., Meng, Q., 2014. Development of a subway operation incident delay model using accelerated failure time approaches. Accid. Anal. Prev. 73 (Issue), 12–19.
Yauch, B., 2014. Toronto Suburban Relief Line. [Online] Available at: http://www.toronto.ca/legdocs/mmis/2014/cc/comm/communicationfile-48033.pdf (accessed 29.07.15.).