A Time-Saving Approach to Simulation Modeling for Traffic Incident Management Program Evaluation

by Chih-Sheng Chou and Elise Miller-Hooks

In this paper, a three-stage time-saving process for conducting traffic incident management (TIM) program benefit evaluation is proposed. This process relies on a developed property-based incident generation (P-BIG) procedure designed to assist in generating a set of incident scenarios that are representative of the historical incident data set and simultaneously not overly large in number so as to induce extensive computational burden. The proposed procedure was applied in evaluating the benefits of an existing TIM program for the purpose of assessing the proposed procedure's predictive power. Results of experiments show that the procedure results in benefit estimates within 5% of the value derived employing all historical incidents, while requiring only 18% of the computational effort.

PROBLEM STATEMENT AND BACKGROUND

Non-recurrent congestion induced by traffic incidents contributes significantly to service level deterioration of both freeways and arterials. A consequence of unstable traffic conditions that result from the primary incident is the occurrence of secondary incidents. Because the occurrence of traffic incidents on freeways and arterials is unavoidable, many traffic incident management (TIM) programs that seek to mitigate the impact of each incident have been widely employed throughout the world. Examples of TIM programs include: freeway service patrol (FSP), automatic incident detection, ramp metering, incident site management, variable/dynamic message sign (VMS/DMS) advisory assistance, route diversion, and professional processing accident scene programs. Such programs aim to mitigate incident impact through quick response, thereby shortening incident duration, or control traffic demand around the incident scene. FSP programs, for example, dispatch patrol trucks along designated beats to detect incidents and assist motorists. Move-it programs encourage or require drivers involved in a minor accident (i.e. with no injuries) to remove vehicles involved in a crash and associated debris out of the roadway (Dunn and Latoski 2003). These programs can be integrated or stand alone.

As states grapple with significant budget deficits, TIM programs around the nation have been the target of cuts. Thus, it has become of increasing importance to show that the benefits of such existing or proposed programs to society significantly outweigh their costs which often include expensive equipment, personnel, overhead, maintenance and publicity. Benefit analyses are used to quantify the social benefits that are derived from improvements in mobility, safety, energy consumption, and environmental impact that result from operating these programs. For example, the benefits of several FSP programs across the nation in terms of travel delay, fuel consumption, secondary incident and emission pollution were estimated (Chang and Shrestha 2000, Chou et al. 2009, Latoski et al. 1999 and Yang et al. 2007). To demonstrate the benefits of controlling traffic demand around an incident by way of a VMS/DMS program in Minnesota, improvements in travel time, total delay and safety that resulted from the program were estimated (Huo and Levinson 2006).

Many studies that seek to quantify the benefits of TIM programs rely on microscopic simulation techniques. Such simulation tools use car-following and lane-changing models to replicate the decisions and movement trajectories of individual vehicles and their response to other vehicles, incidents and geometric design (see May (1990) for addition detail). These techniques offer the

ability to model variability in individual driver behavior, and thus are more flexible than alternative analytical approaches. Moreover, the outcome is often easily understood by experts, as well as the layperson. These studies most often involve two sets of simulation-based experiments and can be categorized as "before and after" or "with and without" studies. "With and without studies" are employed where no "before" program data is available. Given an estimated (or assumed) savings in incident duration of x-minutes as a result of the TIM program implementation, benefits are estimated from two sets of simulation runs: one in which incidents with reported durations are simulated and the other in which the reported durations are extended by x-minutes, replicating the situation where the TIM program has not been implemented. The difference in performance measures between the two sets of runs provides an estimate of savings due to the program. There are shortcomings to either approach, e.g. confounding factors that are difficult to account for in "before and after" studies and a need to surmise what might have been in "with and without" studies.

Although simulation is a popular approach for conducting such benefit studies, it can be quite time-consuming. Thus, several studies report findings based on simulation runs of only a small portion of recorded incidents. For example, Yoshii et al. replicated only a single incident in evaluating the benefits of a dynamic route guidance program (Yoshii et al. 1995). Similarly, only three incidents with different incident durations (22, 26, and 33 minutes) were considered in a series of ITS strategy evaluations involving local and coordinated ramp metering (Chu et al. 2004). In a study of the CHART FSP program in Maryland, 120 incidents out of 1,997 were simulated in estimating the program's benefits (Chang and Shrestha 2000). Because the estimated benefits can vary greatly with the simulated incident properties, the findings may be misleading.

To overcome the shortcomings of simulating only a select portion of recorded incidents, some studies replicate very large numbers of incidents with a wide range of attributes. For example, in analyzing the FIRST program in Minnesota (MNDOT 2004), hundreds of representative incidents were simulated in PARAMICS. The properties of the simulated incidents were carefully defined (with durations between 0 and 40 minutes and varying lane blockage characteristics). These runs resulted in more than 100,000 output files requiring analysis. In another study, 693 historical incidents with distinct incident duration and severity level (by lane blockage) arising along a freeway segment in New York State (all relevant incidents in the historical database) were simulated using CORSIM to evaluate the H.E.L.P. program (Chou et al. 2009). Proper simulation of hundreds of incidents often requires thousands of simulation runs. For example, replication of the 693 historical incidents in studying the benefits of the H.E.L.P. program required 6,930 runs under a single assumption associated with incident duration reduction resulting from the program. While simulation of a large number of incidents with varying properties will produce more accurate benefit estimates, such studies can be quite computationally burdensome, particularly when the number of incidents is large as might be the case where a program covers a wide study area or an area that is densely populated.

The concept of employing randomly generated incidents for the purpose of investigating the benefits of a TIM program within a macroscopic simulation platform was first introduced by Latoski et al. (1999). Such random generation was required in their study to overcome deficiencies in the historical data set. This random generation approach was later expanded in Pal and Sinha (2002) for use in a slightly different context, where the goal was to evaluate various strategies for deploying FSP trucks along roadways in Indiana. The incidents, once generated, were fed into a mesoscopic simulation model that combines microscopic modeling of the FSP trucks with macroscopic models of general traffic. Macroscopic models capture the relationships between flow, speed and density characteristics of traffic flow, and (unlike microscopic models) do not characterize individual vehicle movements. Because the focus of their work was on the evaluation of proposed deployment strategies and a macroscopic approach was employed for traffic modeling, no experiments were conducted to assess whether or not the number of simulation runs could be reduced through the use of such random incident generation. Moreover, few details of the produced incident distributions were provided and only limited assessment of these distributions in terms of how well they represent historical data was completed.

In this paper, the property-based incident generation (P-BIG) procedure is proposed for designing a set of incident scenarios, with incident properties, from historical incident data for use in conducting both "before and after" and "with and without" evaluation studies of existing and proposed TIM programs. This technique can be viewed as a variation on the random incident generation approach conceived in Pal and Sinha (2002). The P-BIG procedure ensures that the carefully selected set of incident scenarios is representative of the historical incident data set and simultaneously not overly large in number so as to induce extensive computational burden. This technique overcomes the deficiencies of prior studies in which either too few, and not necessarily representative, incidents were replicated to ensure valid results or too many incidents were replicated, requiring enormous computational effort and time for output synthesis. Results of this work will benefit police and traffic agencies, especially those in less wealthy jurisdictions, charged with running incident management programs. Procedures developed in this work reduce the effort (and thus cost) for determining whether such a program is worth its cost, or alternatively, defending the program's benefits, potentially saving it from elimination.

METHODOLOGY

A three-stage time-saving analysis process with embedded P-BIG procedure for generating a set of incidents with representative incident properties is presented in this section. This technique involves multiple steps, including incident duration or traffic demand savings estimation, empirically or theoretically derived incident property probability distribution function fitting, scenario generation through randomly generating a small set of incidents from the distributions, simulation running, and results analysis.

In Stage 1 of the proposed three-stage time-saving analysis process for TIM program evaluation, incident and traffic data are collected and analyzed, critical incident property distributions and incident duration savings due to the TIM program are estimated.

In Stage 2, with input from the incident property distributions constructed in Stage 1, a pre-selected number of incidents are randomly generated. That is, for each generated incident, a set of incident properties pertaining to incident severity, type, duration, time of occurrence, and location is generated from the incident property distributions developed in Stage 1. It is hypothesized that, if sufficient in number, these incidents will be representative of the historical incidents and their properties, and likewise, will reflect, in correct proportion, the properties of the historical data. The generated incidents are referred to as the base set, with durations consistent with already implemented TIM programs. A comparison set is generated by appropriately increasing incident duration for each incident by the estimated average savings in incident duration due to the TIM program as found in Stage 1. Note that if a "before and after" study is considered, rather than a "with and without" study, the base set would contain incidents with durations representative of those observed without (i.e. "before") the implementation of the TIM program and the durations associated with the incidents in the comparison set would be reduced appropriately to model the expected savings due to the program (i.e. "after" the program is implemented). In studying VMS/DMS programs, or other programs designed to control traffic demand around an incident, incident durations are constant between the base and comparison sets, and instead, properties associated with prevailing traffic conditions are varied.

In Stage 3, all random incidents within the base and comparison sets are simulated and performance measures are computed. Essential measurements for benefit evaluation are derived from the difference of the pair of measurements from the base and comparison runs.

A flow chart of this three-stage time-saving analysis process for benefit analyses of a TIM program is given in Figure 1. Details of the specific steps associated with this three-stage process are provided in following subsections.

Figure 1: The Three-Stage Time-Saving Analysis Process for TIM Program Benefit Evaluation
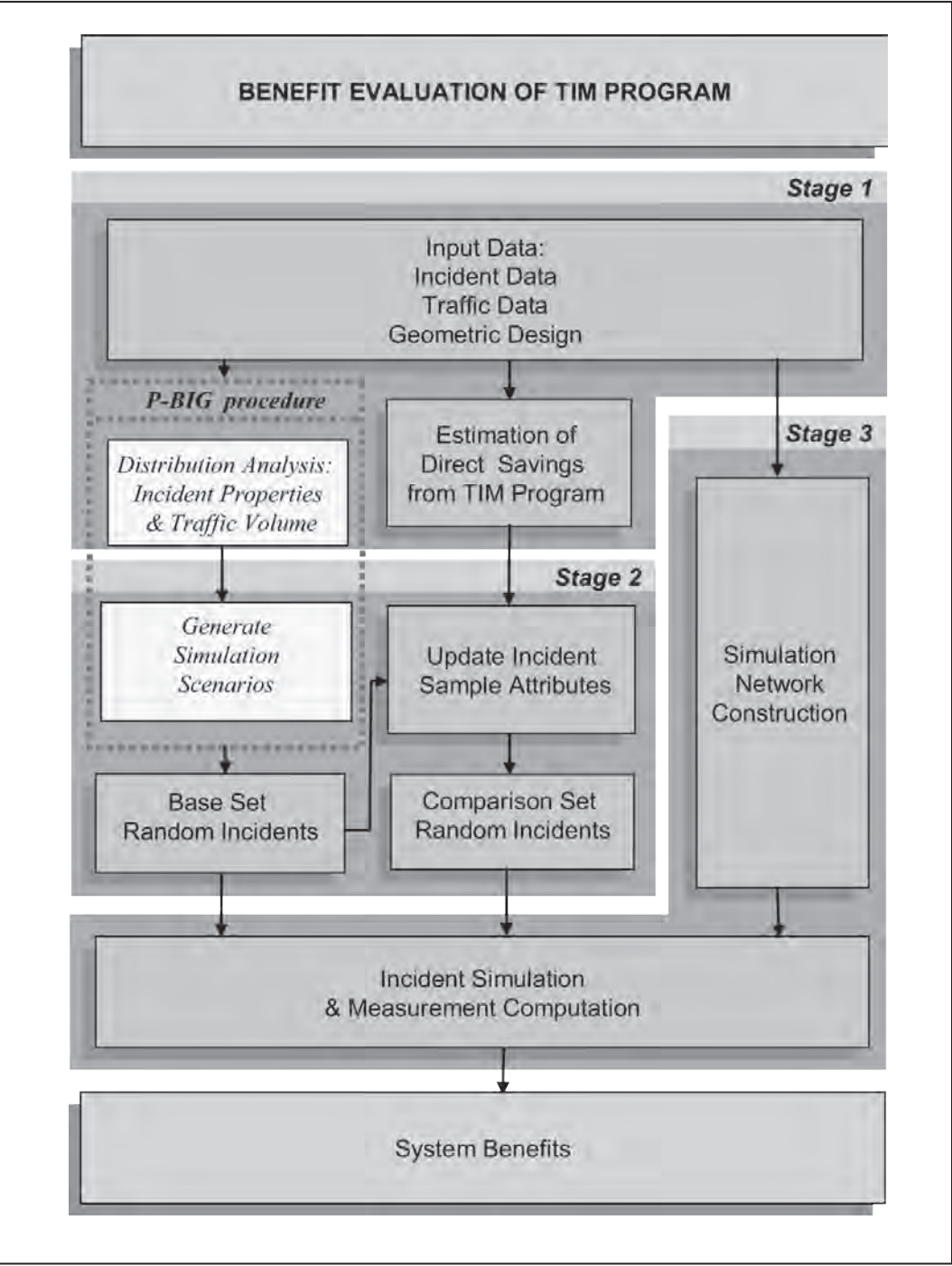
The contributions of this work are derived from the distribution analysis of Stage 1 and generation of simulation scenarios of Stage 2 that together comprise the P-BIG procedure.

Stage 1. The Analysis Stage

To conduct a benefit study of a TIM program, incident data, traffic data, and geometric design associated with the study area must be collected. Once the data are obtained, two main tasks must be conducted in this first stage: 1) estimate direct savings, including reduction in incident duration and/or travel demand, that result from implementation of the TIM program and 2) fit incident property probability distributions. The direct savings in incident duration can be estimated by comparing two groups of incident data sets: "with and without" or "before and after." For example, many FSP program evaluations include "with and without" analyses to estimate average reduction in incident duration that results from implementation of the program. Chou et al. (2009) analyzed a total of 5,508 incidents to which either an FSP personnel or trooper responded. They found that the FSP program saved on average 19 and 20 minutes in incident duration for incidents involving disabled vehicles and collisions, respectively. In addition, reduction in travel demand can be derived from detector reports before and after the implementation of a TIM program aimed at reducing traffic demand around an incident. For instance, Huo and Levinson (2006) compared the detector output for a VMS study and found that approximately 13%-15% of travel demand could be diverted.

The second task of the analysis stage is to fit a probability distribution function for each of the incident property characteristics. These functions are used in generating random incidents and provide an approximation to the historical data. There are several steps for fitting distributions of a sample of incidents with sufficient data points. First, the histogram of incident distributions must be drawn. Certain theoretical distribution functions can be used to fit the shape of the histogram. Specifically, theoretical distributions of exponential, Weibull, log-logistic, gamma and lognormal can be used for fitting incident duration distributions (Nam and Mannering 2000). Once the theoretical distribution is chosen, the parameters associated with the distributions must be estimated. The maximum likelihood estimation method is employed herein for this purpose. For example, the parameter, β, of an exponential distribution, exp (β), can be estimated from the sample mean. Finally, the goodness of fit for a chosen distribution can be tested by computing the chi-square statistic of theoretical and observed frequencies for chosen bins. When no theoretical distribution function is found to match the shape of the histogram, a continuous empirical distribution can be used as shown in equation (1) (Law and Kelton 2000).

$$
(1) \quad F(x) \quad \begin{cases} 0, & \text{if } x < X_{(1)}; \\ \dfrac{i-1}{n-1} + \dfrac{x-X_{(i)}}{(n-1)(X_{(i+1)}-X_{(i)})}, & \text{If } X_{(i)} \leq x < X_{(i+1)} \text{ for } i = 1,2,\ldots, n-1; \\ 1, & \text{if } X_{(n)} < x, \end{cases}
$$

where

x : $x \in \{X_j\}$, $j = 1,2,\ldots n$ incident duration samples;

$X_{(i)}$: i th smallest incident duration sample, i = 1,2,…n

$F(x)$: cumulative distribution function of variable x .

Stage 2. Incident Generation

A preselected number of incidents must be randomly generated. The P-BIG procedure proposed for this purpose is outlined in Figure 2. Incident occurrence is assumed to have a nonstationary Poisson distribution, where incident rates oscillate between high and low frequencies throughout the day. The process proposed herein generates incidents for 24 hours per day. A thinning algorithm that rejects or accepts generated random variates based on time-of-day is employed to produce incidents for select time periods as described in (Lewis and Shedler 1979).

Figure 2: Property-Based Incident Generation (P-BIG) Procedure

Procedure P-BIG

Step 0: i=1; $t_{i-1} = 0$.

Step 1: Create incident I_i with properties $t_i, l_i, m_i, d_i, y_i, s_i, r_i, e_i, v_i$; $t_i = t_{i-1}$.

Step 2: Generate U_1 and U_2 as independent identically distributed $U(0,1)$.

Step 3: Replace t_i by $t_i - (1/\lambda^* \times 24 \times P) \ln U_1$; T=$t_i / 60$.

Step 4: If $U_2 \leq \lambda(T)/\lambda^*$:

    Step 4-1: Return time property, t_i;

    Step 4-2: Generate location property, l_i;

    Step 4-3: Generate incident occurrence month property, m_i;

    Step 4-4: Generate incident direction property, d_i;

    Step 4-5: Generate incident type property, y_i;

    Step 4-6: Generate severity property, s_i, conditioned on incident type;

    Step 4-7: Generate responding unit property, r_i, conditioned on incident type;

    Step 4-8: Generate incident duration properties, e_i, conditioned on incident
            type, lane blockage and responding unit;

    Step 4-9: Assign traffic volume, v_i, to incident based on incident properties and
            traffic data;

    else return to Step 1.

Step 5: If $t \leq 1440$, $i = i + 1$, and return to Step 1; otherwise, stop. The procedure terminates.

Notation employed:

I_i	:	incident sample i; $i \in Z^+$, an integer number for incident sample;
t	:	incident occurrence time (in minutes from midnight), $0 \leq t \leq 1440$;
l	:	mile marker, $0 \leq l \leq L$, where L is the highest mile marker value;
m	:	month, $m \in \{1,2,..,M\}$, where M is the number of months of data;
d	:	direction, $d \in \{E, W, S, N\}$;
y	:	incident type, $y \in \{1,2,...Y\}$, with Y classes of incident type (e.g. collision or disabled vehicle);
s	:	incident severity level, $s \in \{1,2,.. S\}$, with S classes of severity level;
r	:	responding unit, $r \in \{1,2,..R\}$, with R types of responding units (e.g. trooper);
e	:	incident duration, $e > 0$;
v		traffic volume; $v \in Z^+$;
P	:	adjustment factor for controlling number of incidents to be generated, $0 \leq P \leq 1$;
$\lambda(T)$	:	hourly incident rate at the Tth hour, T=(0,1,2...23);
λ^*	:	maximum hourly rate, $\lambda^* = \max\{\lambda(T)\}$.

To apply this procedure, hourly incident rates, $\lambda(T)$, $T = (0,1,2,\ldots23)$, must be computed based on the incident samples. In Step 1, for i, a positive integer, a random incident (I_i) will be assigned with initial occurrence time (t_i), together with properties: location (l_i), month of occurrence (m_i), direction (d_i), incident type (y_i), severity level by lane blockage (s_i), type of responding unit (r_i), incident duration (e_i)and prevailing hourly traffic volume (v_i).

In Step 2, two uniformly distributed random variates are generated. In Step 3, the time of incident occurrence is updated by employing the first random variate (U_1) and the maximum hourly incident rate (λ^*) within the Poisson distribution. Note that P, $0 \leq P \leq 1$, is an adjustment factor to control the number of incidents to be generated. The smaller the value of P, the fewer incidents generated and the fewer the number of simulation runs required. By setting $P = 1$, the number of randomly generated incidents will be approximately equal to the number in the historical incident set. While the accuracy of estimates generated from results of the runs will be improved, the greater the number of incidents considered will result in a trade-off between accuracy and computational effort.

Finally, in Step 4, the generated incident is accepted if the second random variate (U_2) is less than the ratio of its associated hourly incident rate to the maximum rate, $\lambda(T) / \lambda^*$. Once an incident is accepted, the incident duration is set (Step 4-1). Additional incident properties of incident location, month, direction, type, incident severity, responding unit and incident duration are generated in Steps 4-2 to 4-8. The procedure of creating incidents is repeated until a termination criterion based on a bound on t is met. As structured, incidents are generated over a 24 hour period, i.e. if t > 1440, the procedure terminates. If the incident is rejected, no incident properties are generated and the procedure starts over at Step 1.

Location, type, month, and direction associated with each created incident can be directly generated from the appropriate distributions. Incident duration depends on incident type, severity and responding unit. Thus, a conditional distribution is used for generating incident duration once these properties are known (i.e. Steps 4-5 through 4-7). Likewise, severity and responding unit depend on incident type; thus, distributions conditioned on lane type set in Steps 4-5 are employed. This interdependence exists, for example, in the study of FSP programs, where FSP personnel respond to more disabled vehicle incidents than accidents. A final property, traffic volume, v, is assigned to each incident (Step 4-9). To make this assignment, incident properties of time, location, and direction are used to determine the associated traffic volume based on historical traffic data. The user can filter any portion of the generated incidents, such as those occurring only during peak hours and/or only with program involvement.

Note that the procedure proposed herein considers the distributions of the most important properties for reporting incidents in most databases. Greater or fewer properties might be available for consideration, depending on the database used or data collected. The procedure for generating the attributes might be revised accordingly to fit the specific database.

Stage 3. Simulation and Standardization of Measurement of Effectiveness for Comparison

Once the incidents are generated, they can be employed in any simulation model for estimating performance measures. Many commercial microscopic simulation tools, including CORSIM, VISSIM, and PARAMICS, have the feature of modeling incidents. Several performance measures, such as travel delay, fuel consumption, and pollution, are computed from vehicle trajectories of the simulated vehicles recorded in each simulation run. The CORSIM microscopic simulation platform is employed in this study. A benefit of the platform is that incident factors, including onset, clearance, duration, lane closure, capacity reduction caused by rubbernecking effect and warning sign/flair, are readily modeled for any prevailing traffic condition. With respect to modeling traffic incidents, this platform is considered to be more efficient than other microscopic simulation packages (Pulugurtha et al. 2002). Details of the processes required for replicating incidents within the CORSIM simulation platform, including the setting of key parameters, are provided in Chou et al. (2009).

To quantify the benefits of a TIM program, each incident must be replicated twice using different incident properties. The first run uses properties from the base set, while the second run uses properties from the comparison set. Suppose, for example, that a TIM program is estimated to save, on average, 10 minutes in incident duration. Then the only difference between these two runs would be the length of incident duration. An incident in the base set with a duration of 13 minutes would incur 23 minutes when considered as part of the comparison set. By evaluating the impact of the additional incident duration incurred as a result of an incident on average delay, fuel consumption and other measures of importance (in the comparison set) in comparison to the corresponding base set incident impact, one can estimate the benefit of the program savings for the given incident. By summing the benefits of all studied incident pairings (i.e. from base and comparison sets), the total benefits of a TIM program can be estimated. The average daily benefits in terms of savings achieved through incident duration reduction, B_d, over a period, D, can be computed as in Equation 2.

$$(2) \quad \bar{B}_d = \sum_{d \in D} B_d \, / \, n = \left(\sum_{i \in I} \left| P_i^c - P_i^b \right| \right) / \, n \, ,$$

where

D : set of days, d, for which incidents are simulated;

B_d : benefits achieved through incident duration reduction on day d;

n : number of days for running a program, $n=|D|$;

P_i^c : performance measure for incident i simulated from comparison set;

P_i^b : performance measure for incident i simulated from base set;

I : set of simulated incidents, i.

As designed, the proposed three-stage process for TIM program benefit evaluation uses a limited set of incidents whose properties approximate those of the entire historical data set. Savings in computational effort achieved through the proposed method for determining a reduced, but representative, set of incidents for simulation increases with increasing study period length.

NUMERICAL EXPERIMENTS AND CASE STUDY

To assess the proposed three-stage time-saving analysis process for TIM program evaluation, the methodology is tested using data collected over a six-month period (January to June of 2006) for the purpose of evaluating the Highway Emergency Local Patrol (H.E.L.P.) (i.e. a TIM) program in New York State. The H.E.L.P. program runs service patrol vehicles that provide free services, such as changing a tire, supplying a small amount of gasoline, jump starting a battery, pushing a vehicle out of the main lanes and off the freeway, or providing minor mechanical assistance for disabled vehicles. In the case of an accident requiring police or other emergency personnel presence, the H.E.L.P. vehicle driver can call for help and can assist in redirecting traffic around the incident. The H.E.L.P. program is operated along several freeway segments in New York State during the morning (6-10 a m.) and evening (3-7 p m.) peak hours. To assess the proposed methodology in terms of its capability of estimating the program's benefits using only a reduced set of representative incidents, program benefits as estimated by the proposed procedure are compared with program benefits estimated by replicating all incidents occurring in the six month period. Two sets of runs of the proposed methodology were conducted, the first employing approximately 1/12 the number of historical incidents and the second employing 1/6 the number of historical incidents. These two sets

of runs were designed to determine a lower bound on the number of incidents that must be replicated to create a representative set of incidents for procedural implementation. Accuracy of results was also examined with randomly chosen subsets of historical incidents.

Data Details and Distribution Estimation

Six-months of incident data along a 10-mile stretch of I-287, one of the roadways along which the H.E.L.P. program operates, were collected for this study. This roadway segment is located in Westchester County, New York, a New York City suburb. The archived incident data consists of 1,303 incidents, 968 of which occurred during the H.E.L.P. program operational hours. Incident logs describing various properties, including different stages of incident timestamps (start, end, dispatched and arrival times), incident type (disabled vehicle or collision), severity level (number of lanes blocked), direction (east or west), and responding unit (H.E.L.P., trooper or both), are recorded in the database. H.E.L.P. truck drivers responded to 693 of the 1,303 incidents. The average reduction in incident duration because of the implementation of the H.E.L.P. program was estimated at 19 and 20 minutes for incidents involving disabled vehicles and collisions, respectively (Chou et al. 2009). A synopsis of empirical incidents and traffic data is given in this subsection. Properties of incident distributions and results of fitting distributions of incident duration are also shown. Findings from statistical analysis of incident distribution functions were used as input for the P-BIG procedure.

Six-Month Incident Property Distributions. Time-of-day dynamics and the spatial distribution of the 1,303 incidents were analyzed. Higher incident frequencies were observed during the morning and evening peak hours as shown in Figure 3. Incidents occurring during the peak and non-peak hours represent 74% and 26% of all incidents, respectively. It was presumed that traffic flow patterns varied at different times of day, day of month, and location. While the traffic data were not available during the study period (the first half of 2006), average data from the first half of 2007 along the same study roadway segment were available. These data were collected from loop detectors (a traffic surveillance system which records vehicle speed, count and occupancy by measuring change of magnetic field of the detector when a vehicle passes through) at locations depicted in Figure 4. Traffic volume distributions at each of the detector locations for the 2007 traffic data are shown in Figure 5. It was assumed that traffic patterns had distributions in 2006 that were similar to those observed in 2007.

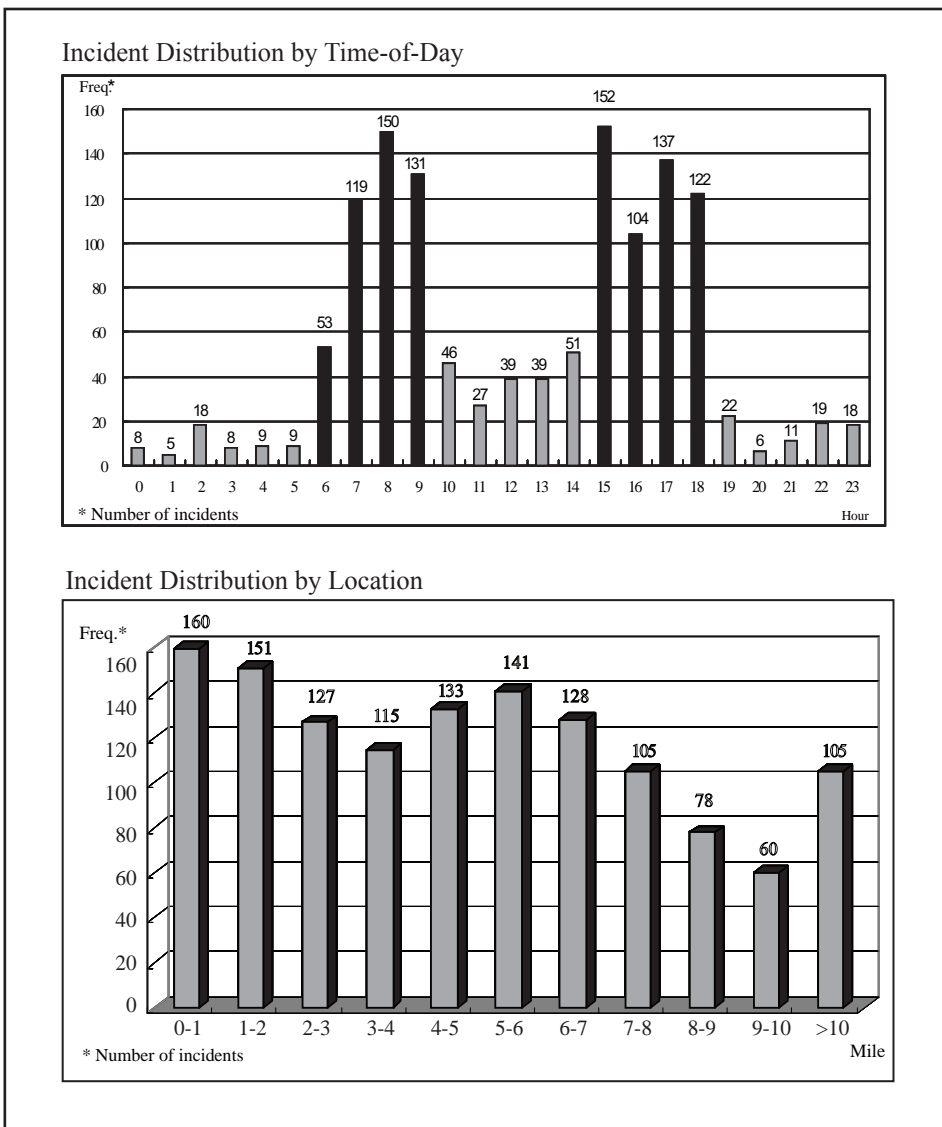**Figure 3: Incident Distributions by Time and Space**

Incident Distribution by Time-of-Day

Incident Distribution by Location
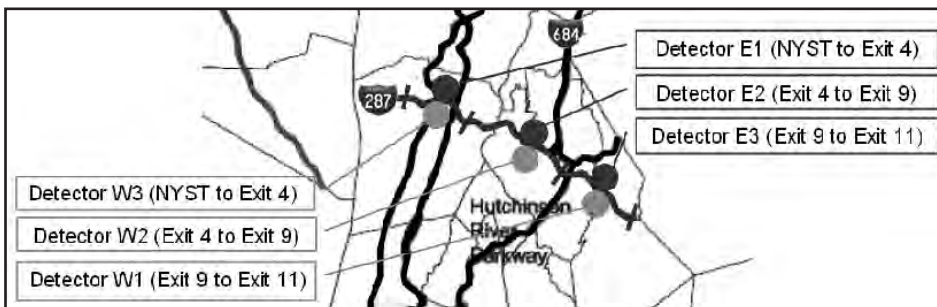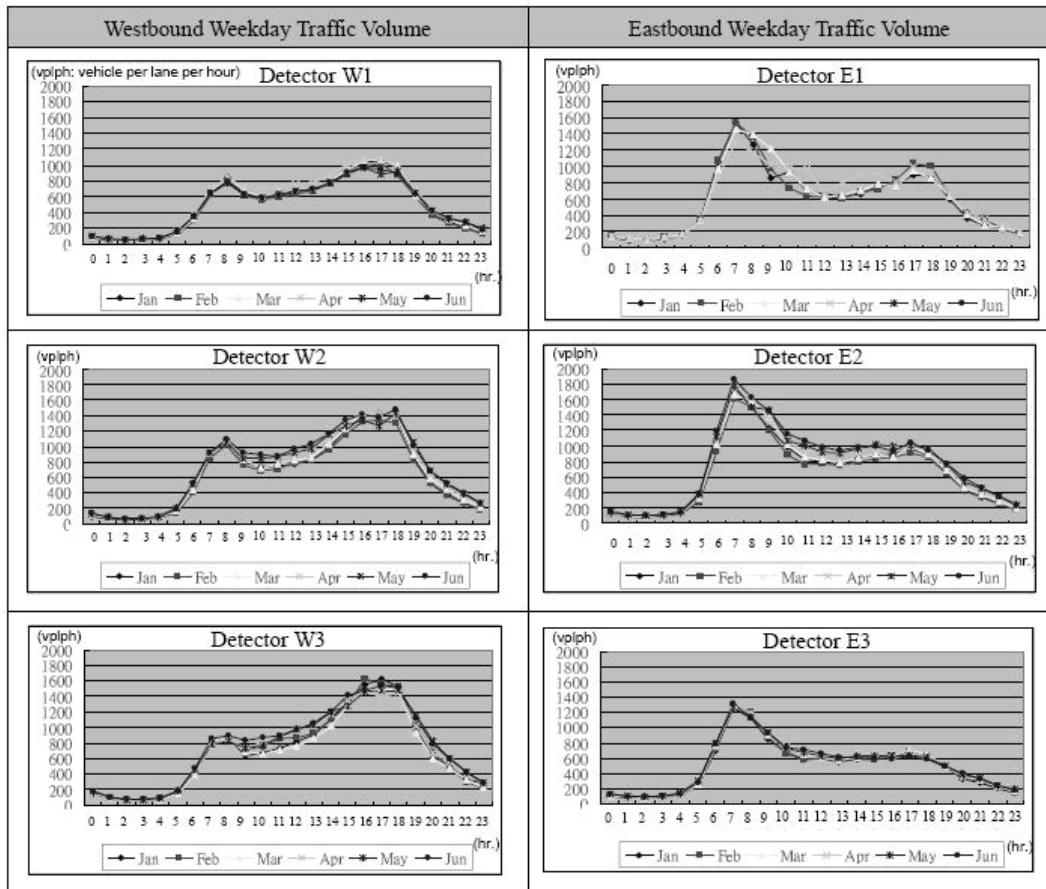

Figure 4: Detector Locations

Figure 5: Time and Space Dynamics of Incident and Traffic Data Distributions

The studied incidents were classified into two categories: disabled vehicles and incidents involving collision (i.e. accidents). During the study period, there were 679 (52%) incidents involving disabled vehicles and 624 (48%) incidents involving collisions reported. The number of lanes blocked by each incident was recorded. The greater the number of lanes blocked, the greater the impact on traffic conditions and the more severe the incident was assumed to be. For the disabled vehicle group of incidents, 91.4% blocked the shoulder. The remaining 8.6% blocked one main lane. For the incidents involving a collision, the shoulder, one lane, two lanes and three lanes were blocked 72.7%, 23.5%, 3.4% and 0.1% of the time, respectively, as depicted in Figure 6.

Figure 6: Incident Distributions by Type and Lane Blockage
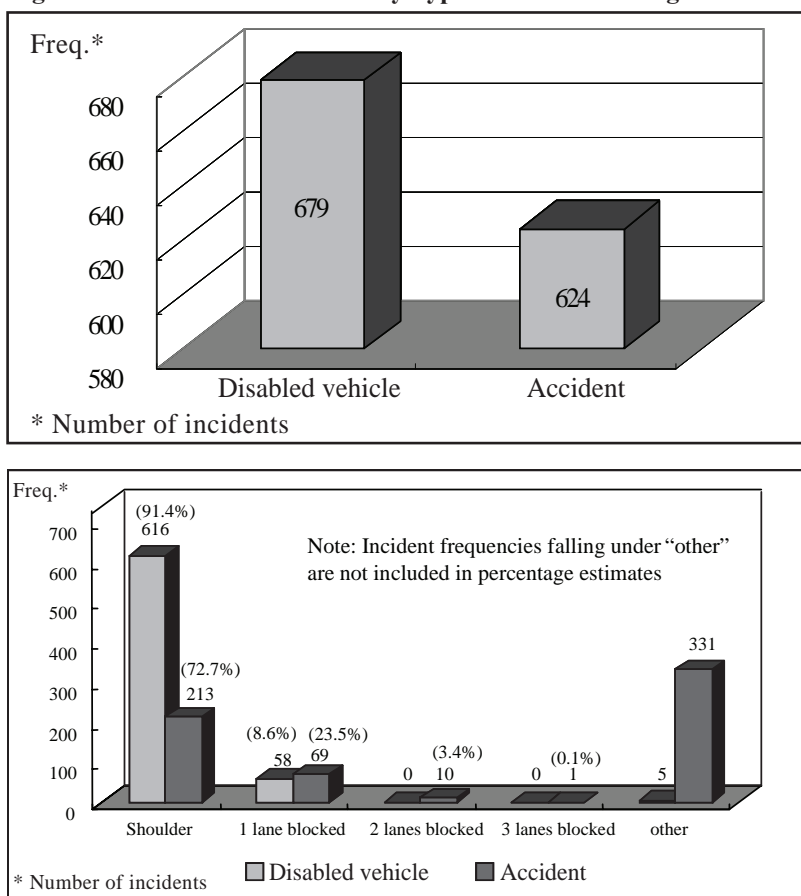
**Probability of Responding Unit Type.** The H.E.L.P. program, like most FSP programs, is designed to assist motorists with disabled vehicles and in collisions involving property damage only. In events involving more severe collisions, i.e. involving injury or fatality, the H.E.L.P. program is designed to provide necessary assistance for the police or to direct upstream (i.e. incoming) traffic safely around the incident scene. Thus, the incidents were classified into "H.E.L.P. only," "Trooper only" and "both H.E.L.P. and Trooper" categories according to their responding unit properties. Only the 917 incidents arising during the H.E.L.P. program operational hours were considered in estimating the probability distribution function of the responding unit type. Incidents classified as "both H.E.L.P. and Trooper" (51 total) were excluded, because no information was available that indicated which responding unit detected the incident first. Table 1 shows the incident types and number of incidents to which either the H.E.L.P. truck drivers or the troopers responded. As indicated in the figure, the H.E.L.P. truck drivers assisted 89% of the disabled vehicles and 24% of the incidents involving collision during the peak hours. By contrast, the troopers handled 11% and 76% of the disabled vehicle incidents and incidents involving a collision, respectively. This information is used in Step 4-7 of the P-BIG procedure provided in Figure 2 to compute the probability that the H.E.L.P. program was involved in a specific incident during peak hours.

Table 1: Incident Response Rates

	Collision	Disabled Vehicle
H.E.L.P. only	78 (24%)	524 (89%)
Trooper only	248 (76%)	67 (11%)
Total	326	591

Incident Duration Distribution for "with H.E.L.P." Incidents. After an incident is generated with its properties and responding unit type in the P-BIG procedure, the incident duration must be specified. Because this duration depends on incident type, severity and responding unit class, conditional probability distributions of incident duration must be developed. For the purposes of the case study, incident durations are only required for those incidents in which the H.E.L.P. program was involved. Thus, all conditional distributions developed in this subsection are conditioned on H.E.L.P. program response.

In estimating these distributions, it was found that, for three of five severity and incident type classifications, the exponential distribution better fits the incident duration distribution than other theoretical distributions, including the lognormal and Weibull distributions, as determined using the Best-Fit software product (Palisade Corporation 2002). The fact that the H.E.L.P. program reported to the database many incidents of short duration may explain why the exponential distribution provides a better fit. The results are shown in Figure 7. Two incident categories, collisions with one and two lanes blocked, are fitted with continuous empirical distributions because no theoretical distribution was found with good fit.

Figure 7: Fitting Incident Duration Distributions using Best Fit
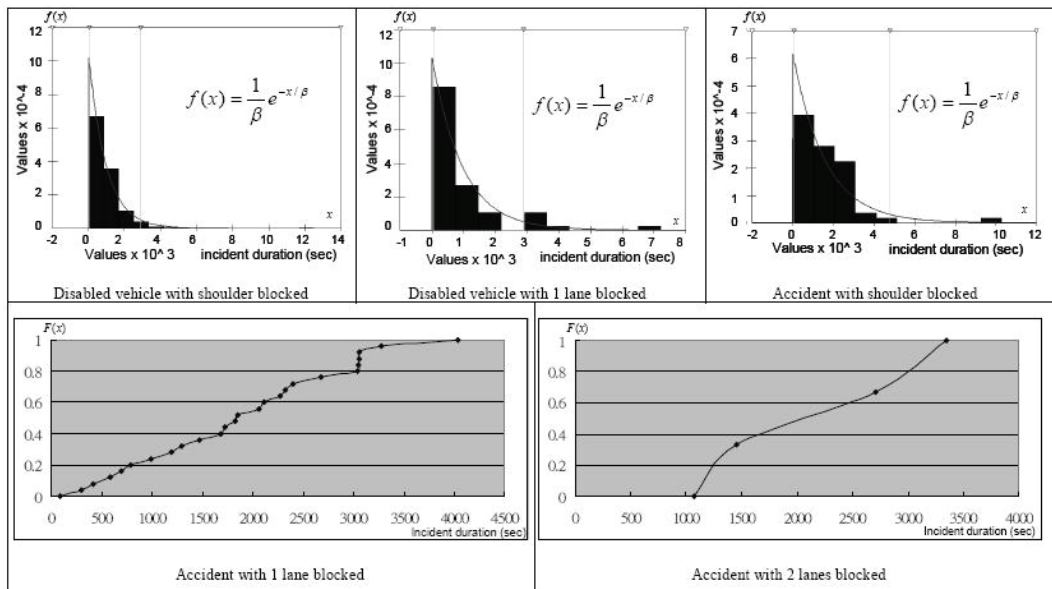
To estimate the parameters of the exponential distributions, the maximum likelihood estimation technique was applied, the results of which are shown in Table 2. The chi-square test was applied to test the goodness of fit of the resulting distributions. It is noted that of the three classes with presumed exponential distribution, only the distribution associated with accidents blocking the shoulder pass this test assuming a 90% confidence level (i.e. with type I error probability $\alpha = 0.10$). While incident duration distributions for incidents involving disabled vehicles did not pass the chi-

square test, the exponential distribution was deemed suitable based on results as displayed in Figure 7 and the fact that no more suitable theoretical distribution could be identified.

Table 2: Results of Incident Distribution Fitting, Parameter Estimation, and Goodness Test

Incident Class	Sample size	Fitting Distribution	Estimated Parameter	Chi-square test result (χ^2 value, $\chi^2_{n,0.9}$, n)
Disabled vehicle with shoulder blocked	507	Exponential	961 (sec)	Fails (48.02, 29.62, n=22)
Disabled vehicle with one lane blocked	52	Exponential	962 (sec)	Fails (16.18, 13.36, n=9)
Accident with shoulder blocked	52	Exponential	1603 (sec)	Passes (11.342,13.36, n=9)
Accident with one lane blocked	26	Empirical	N/A	N/A
Accident with two lanes blocked	4	Empirical	N/A	N/A

Note: "n" is the bin number used for fitting the distributions and chi-square test

Evaluating the P-BIG Procedure

Resulting incident properties from exercising the P-BIG procedure are compared with those of the historical incidents. The following settings were employed within runs of the P-BIG procedure. $Y=2$ (i.e. $y \in \{1(\text{disabled vehicle}), 2(\text{collision})\}$), $M=6$, $d \in \{E,W\}$, $S=3$ (i.e. $s \in \{1(\text{shoulder blocked}), 2(1 \text{ lane blocked}), 3(2 \text{ lanes blocked})\}$), $R=2$ (i.e. $r \in \{1(\text{by H.E.L.P. personnel}), 2(\text{by trooper})\}$) and $P=1$.

Properties of historical incidents were compared with those of a comparable number of incidents generated by the P-BIG procedure for the purpose of evaluating how representative the generated incidents are of the historical incidents. The proposed procedure was applied using a set of randomly chosen seeds (fixing the starting points for the sequence of random numbers used in generating random events), fitted distributions with parameters, and adjustment factor, P, equal to one. Hourly incident rates across different hours of a day were computed. A maximum hourly incident rate of 0.126 was noted to arise during the 8 a.m. to 9 a.m. hour. The comparison between the sample (i.e. incidents generated by the proposed technique) and historical data of incident rates is depicted in Figure 8. As shown in Figure 8, the resulting incident set maintains an incident occurrence rate and distribution over the day that well matches that of the historical data set.

The percentage of incidents to which different units responded in both historical and random incident sets is given in Table 3. It can be seen that the percentage of incidents to which the H.E.L.P. truck drivers and troopers responded are nearly identical for both the historical and sample data sets for both accident and disabled vehicle classes.

Table 3: Comparison of Incident Frequency Percentages by Responding Unit

	Accident		Disabled vehicle	
	Historical	Sample	Historical	Sample
H.E.L.P only	24%	27%	89%	89%
Trooper only	76%	73%	11%	11%

Figure 8: Historical and Random Incident Rates by Time of Day
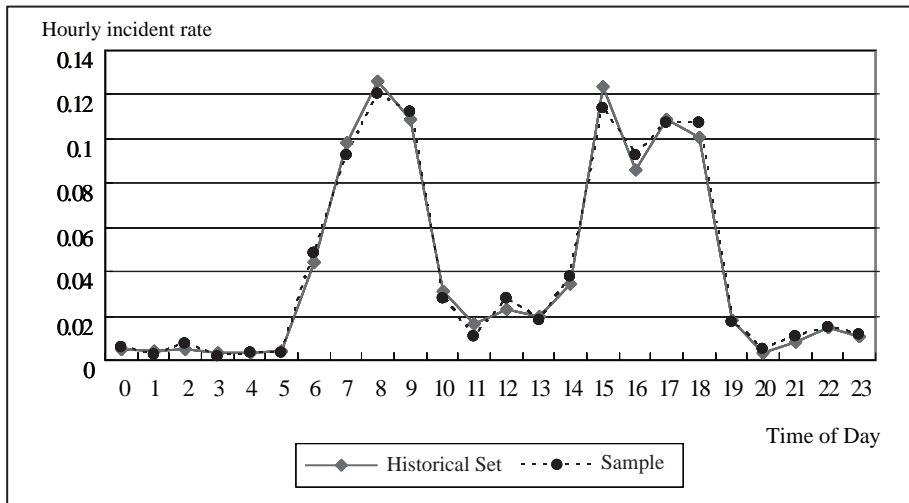
By inspecting the responding unit property of the sample incident set, 688 random incidents (and 693 historical incidents) were identified as having program involvement. Incident duration distributions at 10-minute intervals for these two groups were investigated as depicted in Figure 9. A similar pattern for incident duration is depicted between these two data sets. Likewise, a good match between data sets is noted after conditioning on incident type and severity level (i.e. number of lanes blocked) as shown in Tables 4 and 5 for one of the test sets completed for a given seed.

Figure 9: Comparison of Incident Duration Distribution at 10-Minute Intervals
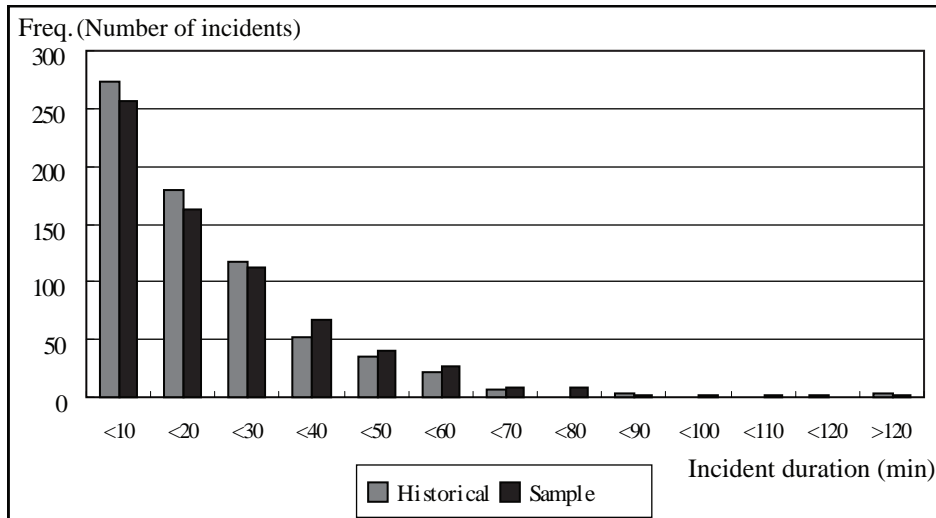
In Table 4, the incident durations are compared by incident type. The average durations for incidents involving disabled vehicles are approximately 16 and 18 minutes for the historical and sample data sets, respectively. For incidents involving accidents, the durations are approximately 29 and 27 minutes, respectively. Not only the values of average incident duration, but also the reported frequencies and standard deviations of the historical and random incident sets, are similar. Resulting severity levels are compared in Table 5. The average duration for incidents with shoulder, one-lane and two-lane blockages ranges from nearly 18-36 minutes and 20-36 minutes for the historical and

sample incident sets, respectively. The frequencies and standard deviations in this table are also similar.

Table 4: Comparison of Incident Duration by Incident Type (minutes)

	Historical			Sample		
	Freq.	Avg.	Stdev.	Freq.	Avg.	Stdev.
Disabled Vehicle	561	16.1	18.8	533	18.1	15.6
Accident	133	28.7	21.7	155	26.8	22.7
Total	693	18.5	20.0	688	20.0	17.8

Table 5: Comparison of Incident Duration by Severity Level (minutes)

	Historical			Sample		
Lane closed	Freq.	Avg.	Stdev.	Freq.	Avg.	Stdev.
Shoulder	601	17.8	19.8	600	19.6	18.0
1 lane blocked	86	22.1	21.0	84	22.0	16.6
2 lanes blocked	6	36.1	13.8	4	35.9	13.4
Total	693	18.5	20.0	688	20.0	17.8

Results of this comparison, thus, indicate that the proposed methodology generates random incidents with historical incident property distributions comparable to those of the original historical data.

In the next subsection, simulation results for varying incident sample sizes are compared with results from runs involving all historical incidents to show that a significant reduction in sample size can produce comparable results to runs on all historical incidents when the proposed P-BIG procedure is employed to generate the set of incidents for sample runs.

Comparison of Simulation Results

The three-stage time-saving analysis process was applied to study the impact on travel delay of the 693 incidents arising over the six-month study period along the 10-mile stretch of I-287. Simulation runs to estimate impact on travel delay were conducted. Specifically, a CORSIM simulation model of the freeway segment with three lanes and one shoulder developed in Chou et al. (2009) was employed. In the previous subsection, it was shown that for $P=1$, the P-BIG procedure produces incidents with similar characteristics to the historical data; hence, one can expect comparable findings in terms of program savings in travel delay if one simulates the random incidents in place of the historical incidents. Computational effort, however, will not be reduced. In this subsection, the impact of testing a smaller number of incidents (generated by the proposed P-BIG procedure) as compared with the number of historical incidents is assessed.

This study first simulated all 693 historical incidents to which the H.E.L.P. program responded (i.e. the *base set*) in the CORSIM model. Given an estimation of 20-minute savings in incident duration due to the program, all incidents were simulated a second time with durations lengthened by 20 minutes (i.e. the *comparison set*). All other factors were assumed to remain constant. Thus, any change in performance is due to the additional delay that results as a consequence of TIM program absence. Five simulation runs for each incident, each with a different seed value, as suggested by Yang et al. (2007) in considering simulation output variability, were conducted. A total of 6,930 replications were, thus, completed. Results of these runs show that an average of 96.4 (or 385.1-288.7) vehicle-hours of travel delay per day were saved due to the H.E.L.P. program. This value is

considered to be "true" and is compared with the results from simulating smaller incident data sets generated by the P-BIG procedure.

Two incident data sets were generated using the P-BIG technique, the first with approximately $1/6^{th}$ ($P=1/6$) and the second with $1/12^{th}$ ($P=1/12$) the number of incidents as compared with the historical data set. Note that for $P=1/6$ the number of incidents generated for the simulation is commensurate with the number of weekdays in a month. Simulation runs of both data sets were conducted, requiring approximately $1/6$ and $1/12^{th}$ the computational effort, respectively.

For $P=1/6$, approximately 120 random incidents were generated and replicated. One hundred twenty replications were completed and savings in average daily travel delay were estimated. To ensure that the results were not specific to any randomly generated set of 120 incidents, the same procedure was repeated 10 times with 10 randomly chosen sample sets and the average daily travel delay savings of the H.E.L.P. program were estimated for each of the 10 sets of runs. A confidence interval was constructed using the student's t-distribution. The performance among these 10 samples shows a 95% confidence interval between 79.9 and 121.3 vehicle-hours of average daily travel delay savings, with an average daily travel delay savings of 100.6 vehicle-hours due to the H.E.L.P. program. Note that the "true" value of 96.4 vehicle-hours falls within the confidence interval. Additionally, the estimated average daily travel delay savings (of 100.6 vehicle-hours) is less than 5% higher than the "true" average daily travel delay savings (of 96.4 vehicle-hours).

This experiment was repeated with $P=1/12$. The 95% confidence interval was constructed, resulting in an interval between 45.4 and 110.6 vehicle-hours, with 78.0 vehicle-hours of average daily travel delay savings. Although the "true" value of 96.4 is also covered within the 95% confidence interval, the average daily travel delay savings (of 78.0 vehicle-hours) is 19% lower than the "true" value (of 96.4 vehicle-hours). The results are displayed in Figure 10. These results indicate that $P=1/6$ provides representative incidents and comparable results, while this is not the case for $P=1/12$.

Figure 10: Confidence Intervals of Simulation Results
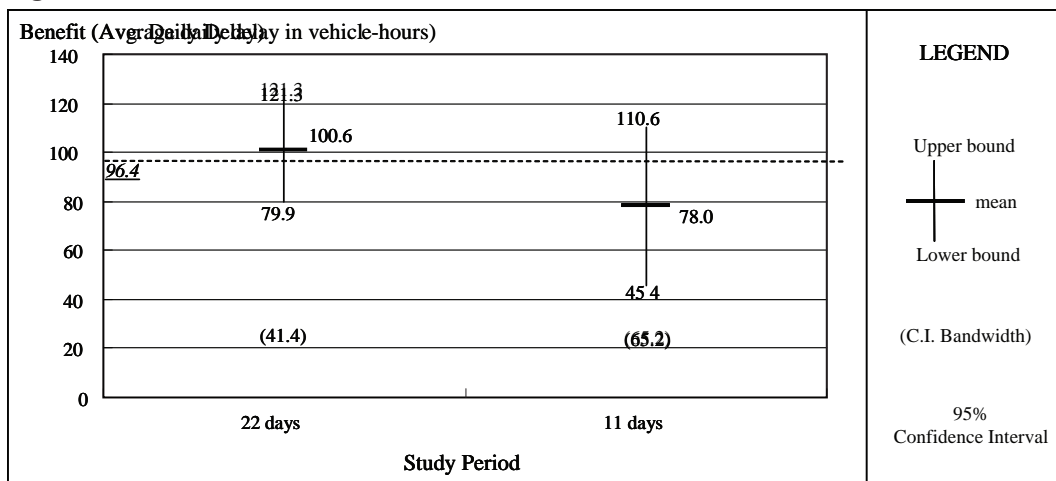
To further assess the P-BIG procedure, results employing the procedure are compared with results gained from simulation of a randomly chosen subset of historical incidents. Specifically, 120 incidents were randomly selected from the 693 historical incident data set and simulation runs of each incident were conducted (and repeated five times for five seed values) for both base and comparison sets. Again, a 20-minute average savings in incident duration due to the H.E.L.P. program was assumed. This process was repeated 10 times and average daily travel delay savings were estimated. The performance among these 10 samples show a 95% confidence interval between 58.9 and 107.7 vehicle-hours of average daily travel delay savings, with mean 83.3 vehicle-hours, due to the H.E.L.P.

program. Note that the "true" value of 96.4 vehicle-hours still falls within the confidence interval. However, the estimated average daily travel delay savings (of 83.3 vehicle-hours) is approximately 14% lower than the "true" average daily travel delay savings (of 96.4 vehicle-hours). Additionally, the confidence interval is significantly wider than the results from the P-BIG procedure as depicted in Figure 11, indicating greater likelihood that the random procedure will provide an erroneous estimate as compared with the P-BIG procedure. In fact, the random procedure results (for 120 incidents) are similar to those of the P-BIG procedure when only approximately 60 incidents are considered. To obtain estimates with a confidence interval of width comparable to that of the P-BIG procedure, nearly 300 incidents would need to be randomly selected as determined in additional experiments. Thus, one can conclude that the P-BIG procedure is beneficial and outperforms simple random incident selection approaches. The P-BIG procedure is estimated to save more than 100%, perhaps as much as 150%, in terms of the number of runs that would be required to obtain equally good estimates if incidents are simply chosen for replication at random.

Figure 11: Comparison of Confidence Interval Results With and Without the P-BIG Procedure
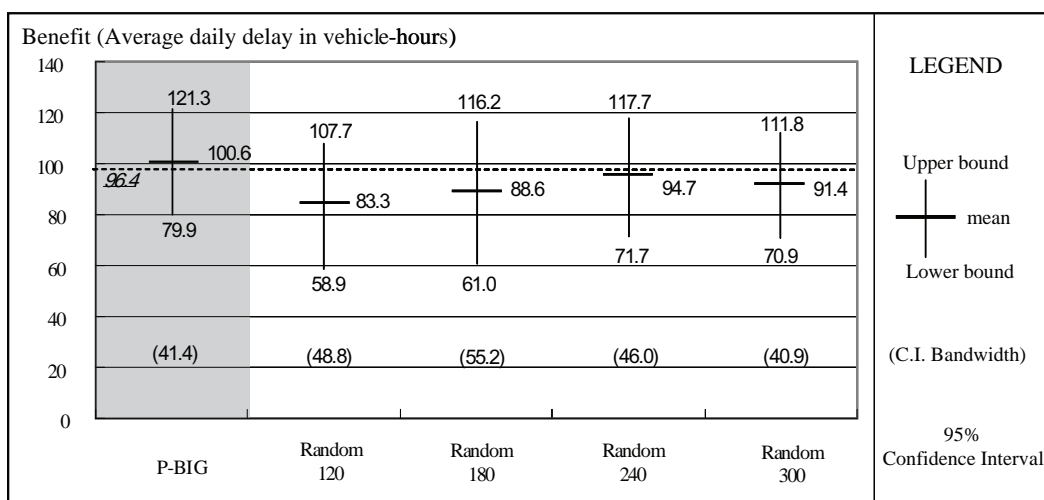
TEMPORAL VARIABILITY AND IMPLICATIONS FOR REDUCING DATA REQUIREMENTS

Data collection and preparation for studies of a TIM program can be quite onerous, regardless of the evaluation methodology used. It was conceived that it might be possible to reduce, not only computational effort required for replication, but also data collection and statistical analyses efforts required for incident property probability distribution fitting and program savings estimation. In fact, a range of time periods (from one month (Yang et al. 2007) to 19 months (Latoski et al. 1999)) for data collection were noted in relevant studies. Thus, it was hoped that one could employ data from only a short time period to fit the incident occurrence and property distribution functions required in the P-BIG procedure. For example, if the incident data show no statistical difference from month to month, then an arbitrary one-month period can be picked to represent important properties of the entire incident data set. Additional experiments using the data collected for evaluation of the H.E.L.P. program as discussed previously were run to assess the viability of employing a reduced data set in generating the distributions employed by the P-BIG procedure. In this section, incident duration distributions across different months are presented and statistically analyzed to determine

whether or not one month of data collection effort could suffice in developing the distribution functions required by the P-BIG procedure.

Incident properties across the six-month study period were considered. Table 6 provides a summary of incident duration by incident severity and type for each month in the study period. It can be seen from this table that incident duration varied significantly across different months for some incident categories. For example, the average duration of incidents involving an accident with two lanes blocked ranged from nearly 18 minutes in April to 56 minutes in June. In addition, there were no such incidents observed in January. Thus, incident data from one month may not adequately represent incident properties for other months of the year. Additional study is required to confirm that the variability is seasonal in nature and not random.

Table 6: Performance of Incident Duration for Different Classes

Incident Class		Month						Total/
		Jan.	Feb.	Mar.	Apr.	May	Jun.	Average
Disabled vehicle with shoulder blocked	Freq:	95	65	77	49	101	126	513
	Mean (min):	13.0	16.5	14.2	13.8	20.4	17.2	16.2
	Stdev (min):	11.8	26.9	14.9	10.6	23.3	17.3	18.6
Disabled vehicle with 1 lane blocked	Freq:	0	27	8	6	3	4	48
	Mean (min):	-	12.5	18.2	13.1	13.5	25.2	16.0
	Stdev (min):	-	24.3	20.6	11.8	11.7	20.2	15.6
Accident with shoulder blocked	Freq:	17	12	8	6	23	21	87
	Mean (min):	27.0	30.9	20.4	32.2	23.5	30.8	27.3
	Stdev (min):	18.9	46.6	18.8	15.3	15.7	21.9	24.0
Accident with 1 lane blocked	Freq:	2	11	12	4	5	5	39
	Mean (min):	56.0	23.6	33.2	28.5	22.1	44.5	31.2
	Stdev (min):	15.8	11.6	17.6	20.9	11.9	11.9	16.8
Accident with 2 lanes blocked	Freq:	0	1	2	1	1	1	6
	Mean (min):	-	34.7	34.7	17.9	38.7	55.7	36.1
	Stdev (min):	-	0	14.8	0	0	0	13.8

A series of Kruskal-Wallis (K-W) statistical tests were applied to test the null hypothesis that each month of incidents comes from the same population (i.e. they have equal populations). The hypothesis is rejected if the K-W H statistic is significant at a test level of 0.05, where the K-W H statistic is computed through Equation (3), assuming that the H statistic follows a chi-square distribution (Kruskal and Wallis 1952). The SPSS statistical software package (Huizingh 2007) was employed to conduct the K-W statistical tests, using $k = 6$, results from which are summarized in Table 7. The hypothesis of equal population was rejected when disabled vehicles blocking the shoulder are considered. This class of incidents is the largest class, involving 74% of all incidents reported in the data collected to study the benefits of the H.E.L.P. program. Thus, using only one month of incident data may not adequately represent conditions over a longer period. Note that the sample size under other incident categories may not be large enough to make a solid conclusion about the sufficiency of employing only one month of data.

$$(3) \quad H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{W_i^2}{n_i} - 3(n+1),$$

where

H : statistic with chi-square distribution with $k - 1$ degrees of freedom;

n_i : number of incidents in month i, $i = (1,2,..k)$; $n = n_1 + n_2 + \ldots + n_k$;

k : number of months considered;

W_i : sum of ranked values for each incident sample in month i.

Table 7: K-W Test Results of Incident Duration Distributions for Different Classes

Incident Class	K-W statistics (chi-square value, P value)	Test result (90%)
Disabled vehicle with shoulder blocked	(16.365, 0.006)	Reject
Disabled vehicle with 1 lane blocked	(5.615, 0.23)	Cannot reject
Accident with shoulder blocked	(4.059, 0.541)	Cannot reject
Accident with 1 lane blocked	(10.481, 0.063)	Cannot reject

Additional experiments were run to assess average daily travel delay savings when replicating all historical data for each month separately. The results for each month are compared with the "true" value of 96.4 vehicle-hours of average daily travel delay estimated from runs replicating all six months of historical incidents. Results of these experiments are given in Table 8.

Table 8: Simulation Results by Simulating Monthly Incident Data Separately

Month	Jan.	Feb.	Mar.	Apr.	May	Jun.	Total
Number of incidents	115	116	107	67	133	156	693
Number of weekdays	20	19	23	20	22	22	126
Total travel delay (vehicle-hours)	336.0	2,196.6	4,460.7	1,896.3	1,011.2	2,245.3	12,146.0
Average daily travel delay (vehicle-hours)	16.8	115.6	193.9	94.8	46.0	102.1	96.4

The results indicate that there is great variability in travel delay savings, ranging from nearly 17 to 194 vehicle-hours saved per day when considering each month separately. Thus, there is a high risk of over- or under-estimating the program's performance with only one month of data. A longer study period is suggested to compensate for short-term variation in incident properties. Such variability may be of more or less significance in other parts of the country. This issue of seasonal variation must be considered when applying any TIM program evaluation methodology on a limited data set.

CONCLUSIONS

The three-stage time-saving analysis process with embedded property-based incident generation (P-BIG) procedure was developed for use in TIM program evaluation in which simulation is applied to assess travel delay savings. The procedure overcomes the drawbacks of approaches applied in existing studies of such programs. For example, some studies experiment with all historical incidents in a study period and, thus, require enormous computational effort, while other studies experiment with only a small subset of randomly chosen incidents from the historical incident dataset. The use of a sample of historical incidents results in significant reduction in computational effort; however, if not chosen carefully, the results of such experiments may over or underestimate program benefits. This study provides a methodology, the P-BIG procedure, for the careful selection of a set of incidents for use in such experiments. The procedure estimates incident property distribution functions based on historical data. These distributions are integrated within a non-stationary Poisson random variate generation process to produce a relatively small set of representative incidents for simulation and derivation of benefit estimates.

To assess the proposed methodology, the three-stage time-saving analysis process was applied on a case study involving a freeway service patrol program in New York State. Six months of empirical data pertaining to the program were examined. Experiments were conducted in a simulation platform in which all historical incidents were replicated, requiring 6,930 simulation runs. Results from these initial experiments showed that an average of 96.4 vehicle-hours of daily travel delay was saved due to the H.E.L.P. program. Additional experiments were conducted on a set of incidents generated by the P-BIG procedure. A savings of 82% in simulation run time and an average error of only 5% were noted as compared with runs involving all historical incidents. When 120 incidents were randomly selected without the assistance of the P-BIG procedure, the average error was more than 14%. To achieve a similar 5% error, nearly 300 randomly chosen historical incidents would need to be considered in the experiments. Thus, careful selection of a set of incidents using the proposed P-BIG procedure results in estimated benefits that nearly perfectly match estimated benefits from runs of all historical incidents with only 18% of the computational effort.

In the case study, monthly variation in incident properties was found to be significant for the six-month study period. This suggests that such variation should be considered in TIM program evaluation studies, as replication of incidents based on properties from only one month could lead to over or underestimation of program benefits. This finding applies not only to the methodology developed herein, but to more traditional simulation-based approaches for studying TIM program benefits.

Additional benefits of the proposed methodology may be derived in benefit studies, where efforts required for data collection are prohibitive. In such circumstances, it may be reasonable to employ the incident property distributions determined in this study, possibly with changes in only the parameters. While imperfect, for many locations and many studies, such input may be sufficient.

References

Chang, G.L. and D. Shrestha. Performance Evaluation of CHART - An Incident Management Program-in 1997. Maryland Department of Transportation, 2000.

Chou, C.S., E. Miller-Hooks, and I. Promisel. "Benefit-Cost Analysis of Freeway Service Patrol Programs: Methodology and Case Study." At press in the *Journal of Advances in Transportation Studies*. (2009).

Chu, L., H. Liu., and W. Recker. "Using Microscopic Simulation to Evaluate Potential Intelligent Transportation System Strategies Under Nonrecurrent Congestion." *Transportation Research Record* 1886, (2004): 76-84.

Dunn, W. and S.P. Latoski. "Safe and Quick Clearance of Traffic Incidents: A Synthesis of Highway Practice." National Cooperative Highway Research Program, Transportation Research Board, 2003.

Huizingh, E. *Applied Statistics with SPSS*. SAGE Publication Ltd., London, 2007.

Huo, H. and D. Levinson. "Effectiveness of VMS Using Empirical Loop Detector Data." California PATH program, Institute of Transportation Studies, University of California, Berkeley, 2006.

Kruskal, W.H. and W.A. Wallis. "Use of Ranks in One-Criterion Variance Analysis." *Journal of the American Statistical Association* 47 (260), (1952):583-621.

Latoski, S.P., R. Pal, and K.C. Sinha. "Cost-Effectiveness Evaluation of Hoosier Helper Freeway Service Patrol." *Journal of Transportation Engineering* 125 (5), (1999): 429-438.

Law, A.M. and W.D. Kelton. *Simulation Modeling and Analysis*. Third Edition, McGraw Hill, Ohio, 2000.

Lewis, P.A.W. and G.S. Shedler. "Simulation of Nonhomogeneours Poisson Process by Thinning." *Naval Research Logistics Quarterly* 26 (3), (1979): 403-413.

May, A.D. *Traffic Flow Fundamentals*. Prentice-Hall, Inc. Englewood, Cliffs, New Jersey, 1990.

MNDOT. Minnesota Department of Transportation – FIRST Program Evaluation. Minnesota Department of Transportation, 2004.

Nam, D. and F. Mannering. "An Exploratory Hazard-Based Analysis of Highway Incident Duration." *Transportation Research Part A* 34 (2), (2000):85-102.

Pal. R. and K.C. Sinha. "Simulation Model for Evaluating and Improving Effectiveness of Freeway Service Patrol Programs." *Journal of Transportation Engineering* 128 (4), (2002): 355-365.

Palisade Corporation. *Guide to Using BestFit – Distribution Fitting for Windows Version 4.5*. Newfield, 2002.

Pulugurtha, S., M. Dangeti, S. Nambisan and M. Kaseko. "Simulating and Analyzing Incidents Using CORSIM and VISSIM Traffic Simulation Software." Conference Proceeding Paper, Applications of Advanced Technologies in Transportation, 2002.

Yang, S., A. Haghani, and D. Iliescu. "A Simulation Approach for Quantifying the Cost-Effectiveness of Freeway Service Patrols Programs." *Journal of World Review of Intermodal Transportation Research* 1 (4), (2007):387-402.

Yoshii, T., H. Akahane, and M. Kuwahara. "An Evaluation of Effects of Dynamic Route Guidance on an Urban Expressway Network." Conference Proceeding Paper, Intelligent Transport Systems World Congress, 1995.

Acknowledgements

Chih-Sheng Chou received his B.S. and M.S. degrees in traffic science from Central Police University, Taiwan, in 1996 and 2001, respectively. He is currently a graduate research assistant and Ph.D. candidate in the Department of Civil and Environmental Engineering at the University of Maryland in College Park. His research interests include freeway operations, incident management, and simulation.

Elise Miller-Hooks is an associate professor of civil and environmental engineering at the University of Maryland. She holds a Ph.D. from the Department of Civil Engineering (1997) and M.S. in engineering (1994) from the University of Texas at Austin. Her research interests are in optimization and mathematical modeling of transportation systems, stochastic and dynamic network algorithms, disaster planning and response, routing and scheduling, transportation infrastructure vulnerability and protection, inter-modal freight transport, hazmat transport, measures for reducing greenhouse gas emissions, concurrent flow lanes, freeway incident management and collaborative and multi-objective decision-making.