

Article

# A Novel Framework for Sustainable Traffic Safety Programs Using the Public as Sensors of Hazardous Road Information

Younshik Chung <sup>1</sup> and Minsu Won <sup>2,\*</sup>

<sup>1</sup> Department of Urban Planning and Engineering, Yeungnam University, Gyeongsan 38541, Korea; tpgist@yu.ac.kr

<sup>2</sup> Department of Civil Engineering, University of Maryland, College Park, MD 20742, USA

\* Correspondence: mswon@umd.edu; Tel.: +1-301-405-2638

Received: 8 October 2018; Accepted: 24 October 2018; Published: 26 October 2018



**Abstract:** Traditionally, traffic safety improvement programs (TSIPs) have been based on the number of crashes at a specific location or their severity. However, the crash datasets used for such programs are obtained from the police and include two limitations: not all crashes are collected by the police (most minor and near-miss crashes are not reported), and the traditional process uses crash data recorded for the past two or three years (meaning most data inevitably include a time lag). To overcome these limitations, this study proposes a new approach for a TSIP based on citizen participation through an online survey that is broadcasted through social media. The method uses the public as sensors of hazardous road information, which means that information can be collected on individual experiences of minor crashes and latent risk factors, such as near misses and traffic conflicts. To demonstrate this approach, a case study was carried out in a small district in the city of Goyang, Korea, which has one of the highest usage rates of social media technologies. The proposed method and a traditional method were both assessed.

**Keywords:** traffic safety improvement program; road risk information; hazardous road location; near miss; empirical Bayes; citizen participation

## 1. Introduction

Various traffic safety improvement programs (TSIPs) and countermeasures have been used at different stages of network development and consider different characteristics in various nations. The first step of a TSIP is the identification of hazardous road locations (HRLs) [1]. Such locations are also called “black spots” or crash-prone locations. HRLs can be defined as any location that has a larger number of crashes than other similar locations as a result of local risk factors [2]. The identification of HRLs requires criteria for what to consider “hazardous.” Several methods have been developed, including the crash frequency method, the crash rate method, the equivalent property damage only (EPDO) method, the rate-quality control method, and various combinations of one or more methods [3].

Although previous methodological approaches are different, they are all based on crash data collected by the police. However, there are two major limitations in using this data. First, not all crashes are reported to the police. When a crash is not critical or there is only minor property damage, most drivers do not usually want to report their crash to police officers for fear of their insurance fees increasing or to avoid the hassle [4]. According to Chung and Eom [5], crash data collected by the police in Korea were estimated to represent less than 20% of the total number of crashes that occur. As a result, the other 80% of crashes are ignored in the TSIP or the identification of hazardous locations, even though the unreported minor crashes could indicate a high risk for injuries or fatal crashes in the

near future [6]. When considering the safety pyramid by Hydén [7], the unreported minor crashes, as well as latent risk factors such as near misses and traffic conflicts, might cause injuries or a major crash [8].

Another limitation is the time lag of crash data. Usually, the TSIP or the HRL identification method uses crash data collected for the previous 3 years. Since the police release the most recent crash statistics in the middle of the next year, the newest crash information could be six months old or more. Therefore, crashes resulting from the same causal factors could occur at the same location in that six-month period. Moreover, the characteristics of the traffic environment, such as the road configuration and the traffic demand, could change from the previous 3 years. As discussed by Behnood and Mannering [9,10], temporal stability might be a major concern in crash-related data. As a result, old datasets could be unreliable for identifying HRLs.

To overcome these limitations, this study proposes a new model for identifying HRLs based on citizen participation (CP). This method uses the public as sensors of hazardous road information through an online survey that is broadcasted through social media in a crowdsourcing approach [11]. Thus, an individual's minor crash information and near misses can be collected. The method can also collect information on latent risk factors, such as near misses and traffic conflicts for roads. To demonstrate this approach, a case study was carried out in a small district in the city of Goyang, Korea, which has one of the highest usage rates of social media technologies. The proposed method and a traditional method were both assessed. The online survey was only broadcasted through social media rather than using social media itself to collect the latent roadway risk information directly.

## 2. Literature Review

### 2.1. Identification of Hazardous Locations

The identification of HRLs is one of the key aspects in the TSIPs of national road safety administrations. Several methods have been developed to identify hazardous locations based on crash histories. They include the crash frequency method [12], the crash rate method [13–15], the equivalent property damage only method (EPDO) [16], the rate-quality control method [17], and the empirical Bayes (EB) method [18–24].

- *Crash frequency*: The simplest method is to use a list of hazardous locations ranked by the numbers of crashes occurring during a given period of time (such as 3 years). Sites are ranked in decreasing order of observed crash frequencies. Since the lengths of the compared segments are different, a modified frequency can be calculated from the total number of crashes over each segment length. This simple method does not take into account traffic volume or crash severity.
- *Crash rate*: This method normalizes the crash frequency with an exposure measure such as traffic volume. It is the most *widely* used method due to its simplicity, but it still does not take into account crash severity. The crash rate of a specific road segment can be simply calculated by dividing the crash frequency in a segment by the segment's length and traffic volume (e.g., average annual daily traffic (AADT)).
- *EPDO*: To combine the frequency and severity score for each crash segment, this method weights crashes by their severity (fatalities, injuries, and property damage only). Each of the injury levels is given a weight that is compared against crashes with property damage only, which are given a weight of 1. This method does not account for exposure.
- *Rate-quality control*: This method was initially proposed to analyze the amount of variability in the crash rate that could be expected as a result of chance for any highway control section [13]. Stokes and Mutabazi [17] developed this method to compare the crash rate at a segment with the average crash *rate* calculated in a group of segments with similar characteristics. Similar segments are assumed to have similar hazard levels. Thus, if the calculated crash rate of a segment is too high compared to similar segments or the average crash rate, that segment is considered to be a hazardous location. This method does not take into account crash severity.

- *EB*: This method was developed to correct for the regression-to-the-mean bias [25]. It combines the crash history of a specific segment with the crash frequency predicted by a crash prediction model. The locations where a high crash count is expected from the EB model can be called HRLs, and a location with the most significant number of crash counts would be ranked the highest.

## 2.2. Citizen Participation in the Transportation Field

CP is a process that enables people to be involved in the planning and implementation of methods to solve problems [26,27]. Some of the important administrative aspects of CP were instituted in the 1960s and 1970s, such as interventions in licensing hearings and broadened standing to sue administrators. However, the basic tools of participation were set in place long before public advocacy was widespread [28]. CP in the planning process may have various significant benefits and lead to better decision outcomes.

The introduction of CP to the transportation field varies by nation. For instance, the U.S. government carried out studies regarding the effects, methods, and guidelines of CP in transportation planning at the beginning of the 1970s [29–31]. With the progression of methodologies and technological environments, many transportation projects have been conducted on involving the public in the transportation planning process (e.g., References [32–35]). Friedman [32] stressed that “planning takes place in a political environment, and that ultimately, all plans are political statements . . . Indeed, all attempts to implement them are political acts.” The U.S. Department of Transportation (DOT) was aware of this political aspect. The Transportation Equity Act for the 21st Century (TEA-21) was enacted on June 9, 1998 and set the policy direction for more comprehensive CP in federal and state transportation planning and decision making. According to TEA-21, the metropolitan transportation planning process includes the following element of public involvement (or CP):

“... A proactive public involvement process that provides complete information, timely public notice, full public access to key decisions, and supports early and continuing involvement of the public in developing plans and transportation improvement programs (TIPs) and meets the requirements and criteria specified”.

As described in TEA-21, CP implies that interaction with citizens should take place during the planning process. Thus, despite the fact that there are many variations of the form, most CP programs in the U.S. have focused on roadway construction planning, new transportation facilities, or technology adoption planning. A number of case studies are available from the Federal Highway Administration (FHWA).

Similar CP approaches were adopted for roadway construction or management programs in other countries. In the early 1990s, Japan’s laws governing urban planning were modified to require more CP in the planning process for local land use and transportation projects [36]. It was mainly implemented for road improvement and management programs at the national level [37,38]. In the United Kingdom, the most visible and memorable moment of CP in transportation planning was associated with major trunk road schemes in the 1990s. There was a public inquiry for the Salisbury bypass, where the DOT’s proposals were ultimately rejected, as well as complaints by environmental pressure groups to the European Commission regarding the Government’s handling of the Environmental Impact Assessment process at Oxleas Wood. There was also a series of protests along the routes of trunk roads, including the historical conflicts at Twyford Down and Newbury [39]. A few studies have conceptualized and implemented CP in national and local transportation planning in the United Kingdom (e.g., References [39–42]). Unlike other countries, CP in Korean transportation projects was reviewed prior to the 1990s, but most studies have been limited to suggesting policies and their necessity [43–46].

Previous studies have indicated that some of the keys to the success of CP efforts are the serious and timely treatment and follow-up of citizen input, such as online comments, incoming calls, and emails [47,48]. Therefore, the traditional techniques of CP, such as personal interviews, public meetings or hearings, and telephone techniques, could rarely be successful. However, recent advances in information technologies hold great potential for the success of CP programs, such as web-based geographic information systems (web-GIS), geo-tagging, social media, and mashup. For example, many people cannot attend public meetings due to scheduling conflicts and limited time. However, online tools enable the public to maintain involvement and issue opinions in their available time. Additionally, some people are not comfortable expressing their opinions in public, particularly if they feel they are in the minority. However, they might feel better able to express themselves in an online environment.

### 2.3. Literature Review Summary

Since there is no universally suggested approach for identifying HRLs, several methods have been developed based on crash histories to reduce identification bias. Most previous studies do not describe the availability and reliability of crash data and assume that the crash data are complete. However, the crash data used in HRL identification are collected by the police and for only 20% of the total crashes at most. Therefore, regardless of the various methodological approaches to identify HRLs, a study is warranted to examine identification based on unreported latent crash risk factors, including minor crashes and latent risk factors.

Unreported crash data can be collected through CP programs. In the past, serious efforts and time were needed to promptly address citizen input. However, recent advances in technologies hold great potential for CP programs. As a result, a CP program based on social media will be able to collect information on individuals' unreported minor crashes and latent crash risk factors. Using such information is a challenge in research on identifying HRLs.

## 3. Identification of HRLs Based on Citizen Participation

### 3.1. Citizens as Sensors of Latent Roadway Risk Information

The increasing popularity of the World Wide Web since the mid-1990s has brought people together by creating virtual communities in cyberspace [49,50]. People can use it to create, receive, share, and exchange information. In recent years, computer-mediated tools have been introduced for virtual communities and networks. These social media have three key elements: content, communities, and Web 2.0 [51]. Content includes photos, videos, status updates, tags, links, and playlists that people create and share. Communities provide opportunities for people to communicate, network, and collaborate, while Web 2.0 refers to easy-to-use technologies and applications that make communication and content sharing possible for ordinary people [52].

Social media technologies challenge the traditional methods of CP by enabling more direct, real-time, and networked ways of acting. Recent studies have used social media for various purposes. Examples include studying the relationships between happiness and mobility patterns [53], tourist origins and attractions [54], and disaster-related behaviors, such as user intentions, information generation and dissemination, and user interactions on social media [55–60]. These studies use people as “sensors”. In this concept, people act as non-technical sensors with contextual intelligence and comprehensive knowledge [61].

### 3.2. Social Media for Citizen Participation

There are many similar definitions of social media. Kaplan and Haenlein [62] defined it as a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0. They allow the creation and exchange of user-generated content, which includes blogs, wikis, discussion forums, posts, chats, tweets, podcasting, pins, digital images, video, audio files, and other forms of media created by users of an online system or service. Social media is the collective of online communication channels dedicated to community-based input, interaction, content sharing, and collaboration. Thus, the advent of social media has revolutionized public relations, information exchanges, and communication. As social media becomes more pervasive, it can be useful for identifying social issues and information.

Recently, the VTT Technical Research Centre of Finland conducted a research project on the possibility of collaboration between citizens and public agencies through social media [63]. They described five benefits of social media tools: (1) the possibility of participating regardless of time and space; (2) continuous communication between face-to-face meetings; (3) follow-up activities after meetings; (4) documentation of discussions and decisions serving as reminders of project details, and (5) automatic archiving of the discussed material. Based on these benefits, social media has the potential for the continuous collection of roadway risk information and providing such information to road users in real time. It could thus contribute to the proactive prevention of traffic crashes. Therefore, to assess the feasibility for identifying HRLs based on CP through social media, this study demonstrates a new program to collect roadway risk information and a new model for identifying HRLs based on an online survey that is broadcasted through social media.

## 4. Case Study

### 4.1. Roadway Risk Information Collection

The roadway risk information collected by CP includes participant characteristics such as age group and gender, as well as the time period, weather conditions, and location of experienced crashes or latent roadway risk. To observe geo-spatial data regarding the location of latent roadway risk information, a simple location-tag function was added to a social media application that is successfully being operated by the city of Goyang, Korea ([www.facebook.com/goyangcity](http://www.facebook.com/goyangcity)). The operators of this application advertised about the purpose of this study to the residents of Goyang, which attracted participants. We added a special location-tagging function to identify the location of the latent roadway risks. The residents then reported the latent roadway risk factors, such as the location and time of hazardous roadways and their personal experiences with near misses or minor traffic crashes that were unreported to the police. In this study, a “near miss” is defined as an incident that did not result in a crash but had the potential to do so.

Since this program was only being tested for feasibility, it was operated for just one week and for an area of about 9 km<sup>2</sup>. To enhance the program participation and the quality of information from the participants, we limited the contributions to one piece of risk information per participant and offered them a gift card. It was assumed that the latent roadway risk information submitted by the residents would be related to recent experiences rather than their old memories. Therefore, we did not limit the time period for their experiences.

In spite of the short advertising and operating periods for the program, a total of 444 items of latent roadway risk information were obtained from 23–29 June 2013. Table 1 shows descriptive statistics extracted from the participants for the identification of HRLs. Since the location-tagging function was added, the location information for hazardous roadways can be compared to other risk information shared by residents.

**Table 1.** Descriptive statistics extracted from participants for the identification of hazardous road locations (HRLs).

		Samples	Percentage
Total		444	100
Gender	Male	192	43.2
	Female	252	56.8
Age	10s	26	5.9
	20s	270	60.8
	30s	99	22.3
	40s	30	6.8
	50s	15	3.4
	60s or older	4	0.9
Time period	Morning (06:00–10:00)	42	9.5
	Daytime (10:00–17:00)	137	30.9
	Evening (17:00–21:00)	132	29.7
	Night (21:00–06:00)	39	8.8
	Always	60	13.5
	n/a	34	7.7
Weather conditions	Clear	185	41.7
	Cloudy	67	15.1
	Rainy	48	10.8
	Snowy	5	1.1
	Foggy	6	1.4
	Always	78	17.6
	n/a	55	12.4

## 4.2. Feasibility Assessment of the Case Study

### 4.2.1. Assessment Framework and Dataset

The proposed method was assessed by comparison with a traditional method. Each set of HRLs identified by the two approaches can be assessed by comparing them with real crash data during the next one-year period. The EB method was selected for comparison since it is one of the state-of-the-art approaches to identify HRLs. Figure 1 shows the framework for the feasibility assessment of the case study, which has the following steps:

- *Arrangement of two datasets:* Two datasets were used for the identification of HRLs: one composed of crash data from three years (2010 to 2012), and another composed of resident-reported latent roadway risk information from 23–29 June 2013.
- *Identification of potential HRLs using the EB method:* The two different sets of potential HRLs were ranked in decreasing order of crash frequency or latent risk frequency with respect to the AADT of each segment.
- *Identification of real HRLs using the EB method:* Since the potential HRLs from the two datasets were identified using datasets prior to 1 July 2013, the real HRLs were identified based on crashes that occurred during a half-year period from 1 July 2013 to 31 December 2013, using the EB method.
- *Assessment of matching rates:* The two lists of potential and identified real HRLs were compared based on how much the lists matched the real HRLs.

We collected crash data from the police, which included a total of 1028 crashes for the study area from 2010 to 2012. Other police-recorded data were obtained as reference points for the half-year period from 1 July 2013 to 31 December 2013, which included 201 crashes.

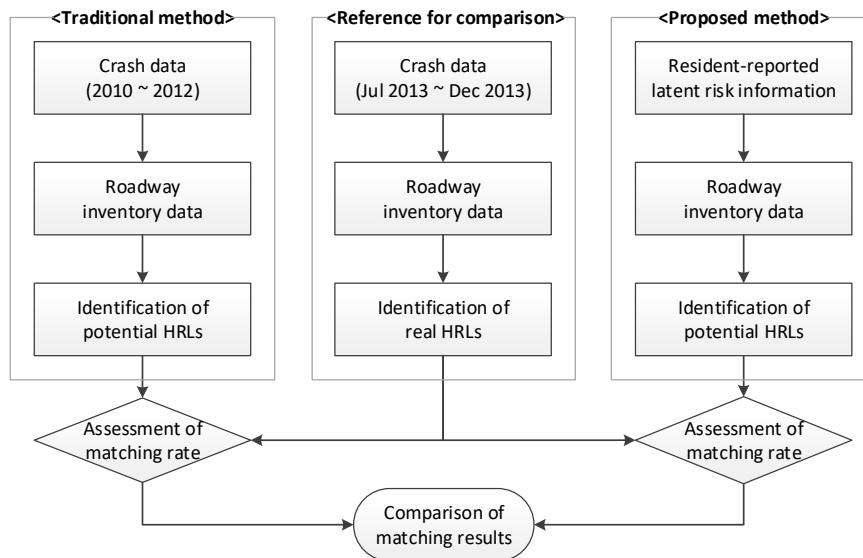


Figure 1. Framework for the feasibility assessment of the case study.

4.2.2. Arrangement of Road Network and Risk Dataset

The sections of a road network must be classified to identify HRLs. Each of these sections is assumed to have homogeneous traffic characteristics, such as traffic volume, geometric roadway conditions, and other factors that influence crashes [2]. As described by Elvik [2], the classification method varies for each country. This study applied a method that is used in Korea [64]. Based on this method, urban areas are basically composed of intersections and roadway segments. With each intersection as a center, the roadway sections within 50 m from the centers are defined as “intersections”, and the other sections outside this range are defined as “roadway segments”. The roadway segments are divided further when they are longer than 200 m. In addition, the intersections are also divided further when the distance between two consecutive intersection centers is less than 100 m. In such cases, the intersection radius could be less than 50 m. Figure 2 shows the results of the road network section splitting for the study area using a generic GIS tool.

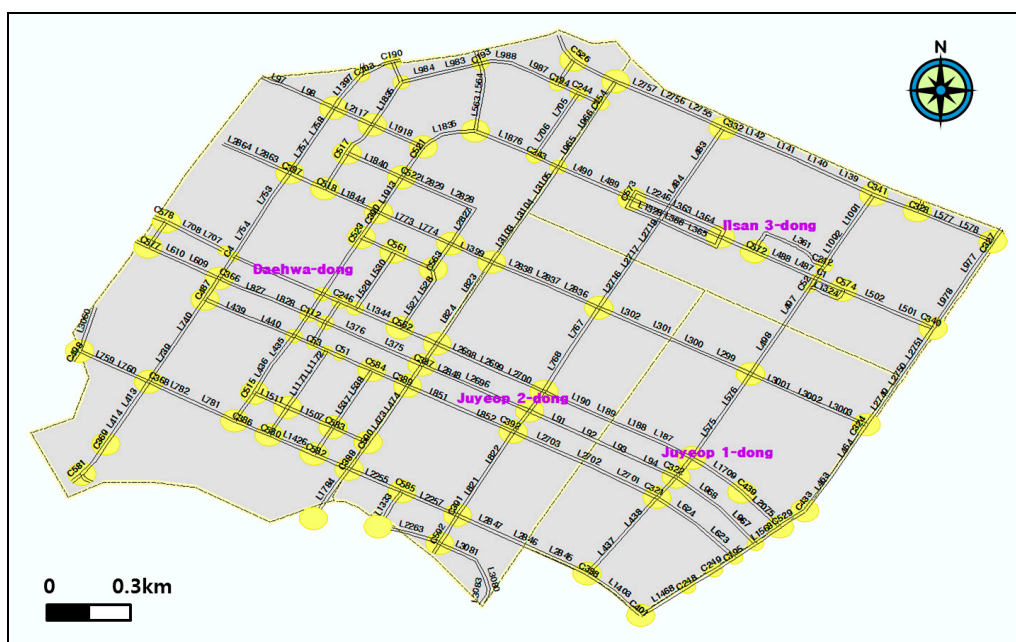


Figure 2. Result of road network splitting to identify HRLs in the study area.

Each road risk dataset can be plotted on a split road network. The crash locations from the past 3 years are shown in Figure 3, and the latent risk locations observed by residents are shown in Figure 4. Some of the sections have zero crashes, but others have one or more. A priority ranking was produced for each dataset. Figure 5 shows the crash locations plotted on the split road network in the study area.

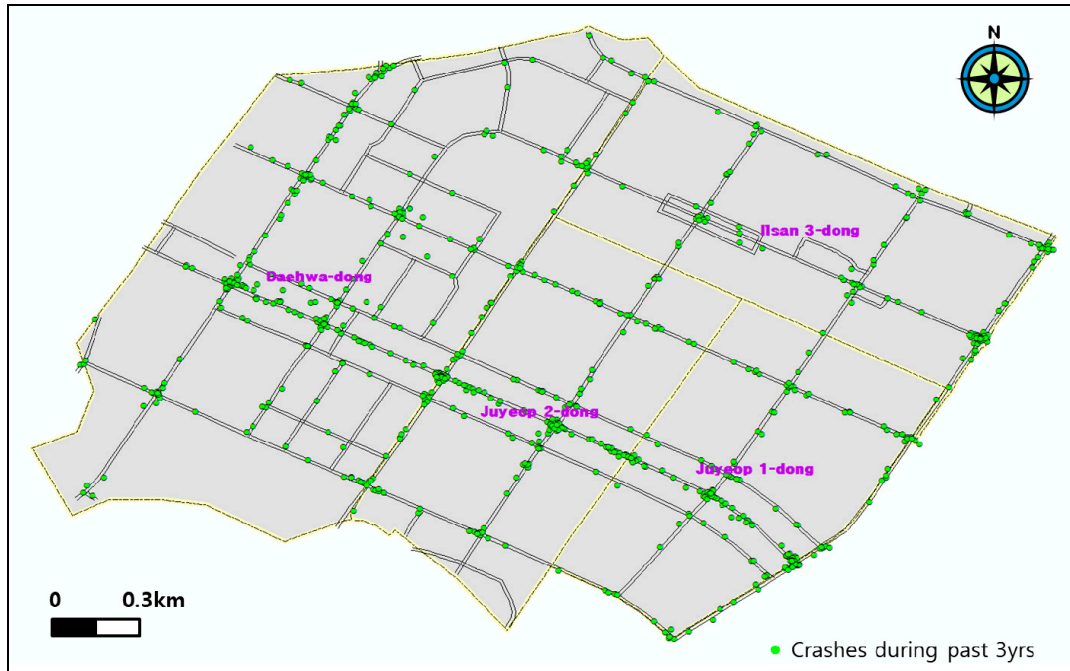


Figure 3. Crash locations from the past 3 years in the study area.

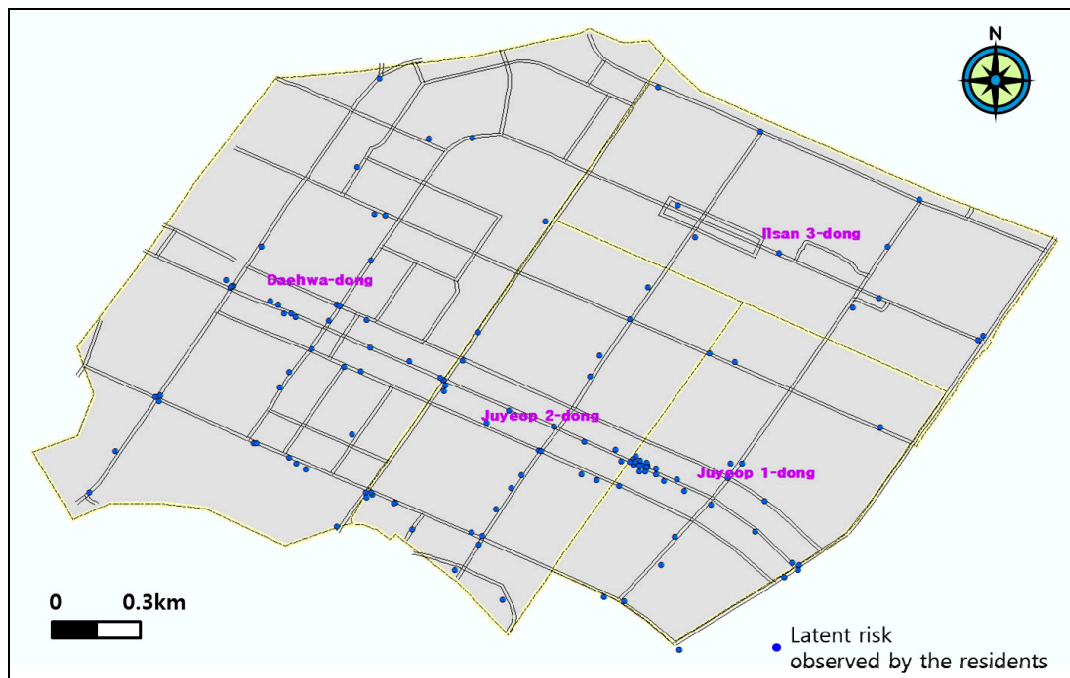


Figure 4. Latent risk locations posted by residents in the study area.



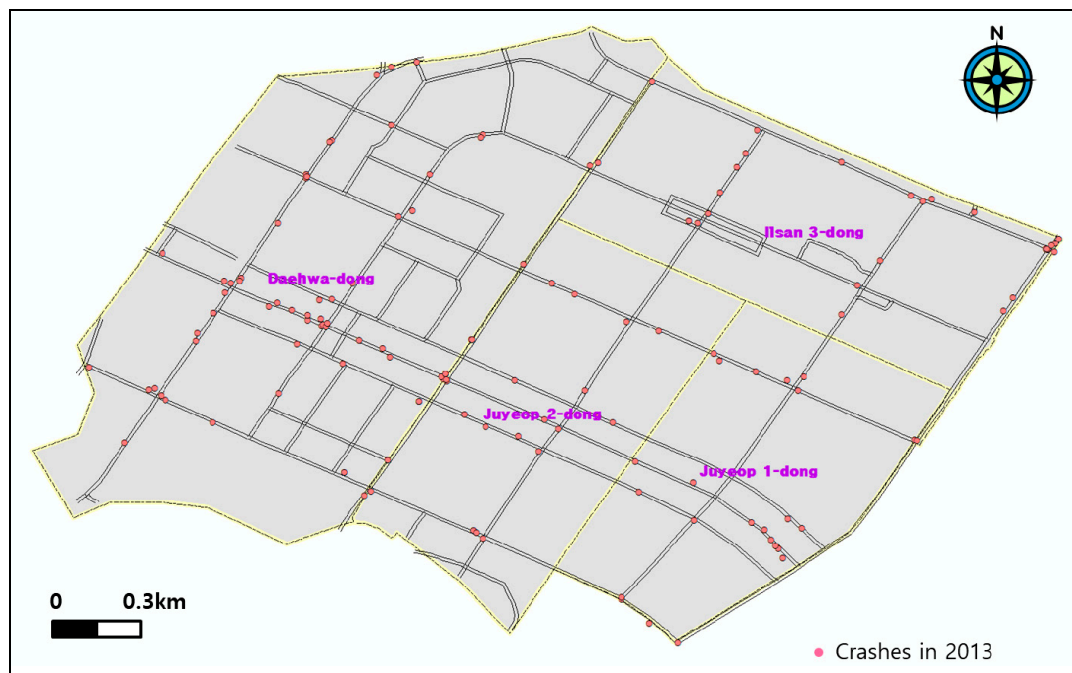


Figure 5. Crash locations that occurred during a reference year in the study area.

#### 4.2.3. Assessment of the Proposed Model

##### Application of the EB Method for Identifying HRLs

The EB method is a popular analysis method for the identification of HRLs for roadway sections. This method can overcome the limitations of other approaches by accounting for regression-to-the-mean effects, traffic volume changes, and time trends in crash occurrence due to changes over time in factors such as weather, crash reporting practices, and driving behaviors [65]. In general, the EB-adjusted crash counts ( $\lambda_i$ ) on roadway section or location  $i$  can be estimated using the following equation:

$$\lambda_i = \alpha_i \cdot E\{\lambda_i\} + (1 - \alpha_i) \cdot x_i \quad (1)$$

where  $i$  denotes a roadway section,  $\lambda_i$  is an EB-adjusted crash count,  $E\{\lambda_i\}$  is the expected crash counts on a roadway section with similar road geometry characteristics or traffic characteristics,  $x_i$  is the crash count on each section, and  $\alpha_i$  is a weight that is estimated by:

$$\alpha_i = \frac{1}{1 + \frac{VAR\{\lambda_i\}}{E\{\lambda_i\}}} \quad (2)$$

The expected crash counts on similar roadway sections can be determined by the safety performance function (SPF), and the weight is between 0 and 1. If the SPF assumes a negative binomial distribution, an overdispersion parameter can be used to determine the weight. The overdispersion parameter shows the relation between the mean and variance, and its significance indicates that the underlying probability distribution is correctly identified as negative binomial [66]. This overdispersion parameter can be adjusted by the characteristics of each individual roadway segment, such as the lengths of each segment [66]. Therefore, the weight ( $\alpha_i$ ) considering a segment-specific overdispersion parameter ( $\phi_i$ ) can be estimated using the following equation if the SPF has a negative binomial distribution:

$$\alpha_i = \frac{1}{1 + \frac{E\{\lambda_i\}}{\phi_i}} \quad (3)$$

The result is determined by how much “weight” is given to the crashes expected on similar roadway sections [24]. The strength of the EB method is in the use of a weight that is based on sound logic and real data [24].

Prior to identifying hazardous road locations using the EB method, the SPFs must be developed for the estimation of the expected crash counts for each type of roadway. Based on three years of crash data (2010–2012), two SPFs were developed for intersections and roadway segments. Historically, crash frequencies have often been assumed to follow a Poisson distribution [66]. However, as shown in Table 2, the variance of the crash counts exceeds their average. This violates the constraints of a Poisson distribution. Therefore, a negative binomial distribution is used to represent the distribution of the crash counts in this study [67].

**Table 2.** Descriptive statistics of different roadway types in years 2010–2012.

	Intersections (80 Sections)		Roadway Segments (286 Sections)	
	Average	Standard Deviation	Average	Standard Deviation
Number of crashes	7.96	11.46	1.91	3.00
AADT	11,719	99,667	12,952	13,486
Length of segment (m)	43.96	11.32	121.01	35.40
Number of lanes	4.39	1.91	4.25	1.85
Speed limit	56.75	16.37	54.67	17.35

Table 3 shows the estimated models for different types of roadways. AADT and the length of each segment are significant in the two models, and the number of lanes is significant for the model for the intersections. The weights and the EB-adjusted crash counts can be obtained using these estimated SPFs and overdispersion parameters, and we can identify the HRLs for the study area.

**Table 3.** Estimated safety performance functions (SPFs) for different types of roadways.

	Intersection			Roadway Segment		
	Estimate	Std. Error	<i>p</i> -Value	Estimate	Std. Error	<i>p</i> -Value
(intercept)	−0.3370	0.4020	0.4019	−0.1531	0.2275	0.5010
AADT/10,000	0.6415	0.0882	<0.0000	0.2015	0.0463	<0.0000
Length of segment	0.0160	0.0079	0.0417	0.0040	0.0018	0.0306
Number of lanes	0.1648	0.0460	0.0003			
	Overdispersion parameter: 0.6619			Overdispersion parameter: 0.3618		

#### Comparison and Discussions for Identified HRL Results by the Traditional and Proposed Methods

The rankings of the HRLs from each dataset were used to assess the potential of CP-based latent risk information and not the number of crashes for each segment. The reason is that the number of crashes for each segment from the resident-reported latent risk information depends entirely on the response rates and the experiences and memories of the respondents, so it is difficult to specify the term of the collected information, and the scales may also differ. Table 4 and Figure 6 show the HRLs from the two identification methods and their comparison results.

**Table 4.** Comparison of identified HRLs between the traditional and proposed method.

HRLs in the Reference Year		HRLs			
		Past 3 Years of Data cf. Reference Year Data		Data by CP cf. Reference Year Data	
% of High Ranked Sections	Number of Sections	Number of Identified Sections	Identified Rate *	Number of Identified Sections	Identified Rate
1% ≤	4	3	75%	4	100%
3% ≤	11 (4 + 7)	9	82%	9	82%
5% ≤	18 (11 + 7)	15	83%	17	94%
10% ≤	37 (18 + 19)	34	92%	36	97%
20% ≤	73 (37 + 36)	68	93%	70	96%
50% ≤	183 (73 + 110)	179	98%	178	97%
80% ≤	292 (183 + 109)	287	98%	288	99%
100% ≤	365 (292 + 73)	365	100%	365	100%

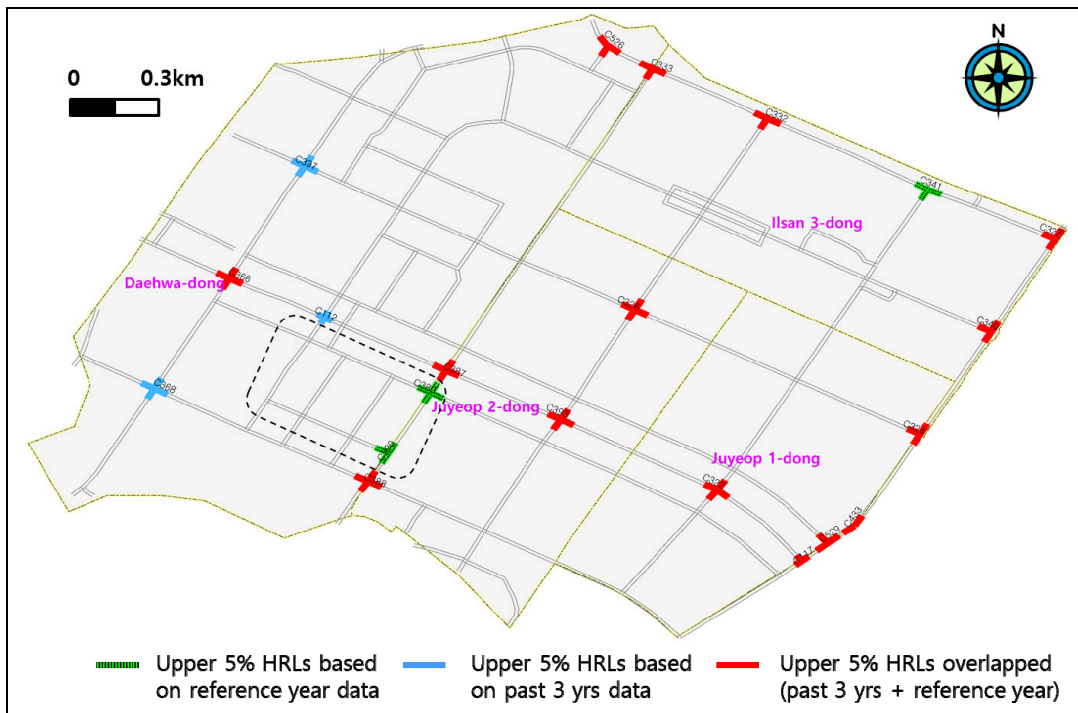
\* Calculated as the number of sections divided by the number of identified sections.

Table 4 includes the comparison results of the identified hazardous locations between the traditional method using the police-reported information and the proposed method using the resident-reported information. The HRLs in the reference year are listed in decreasing order of high-risk sections with all of the sections in the study area. Since this study applies the EB method, a section that is associated with more crashes is likely to be a riskier section. Some sections could have tied rankings. As shown in the results, the proposed method using the resident-reported information is a little bit better than the traditional method using the police-reported information for the overall high-ranked risk sections.

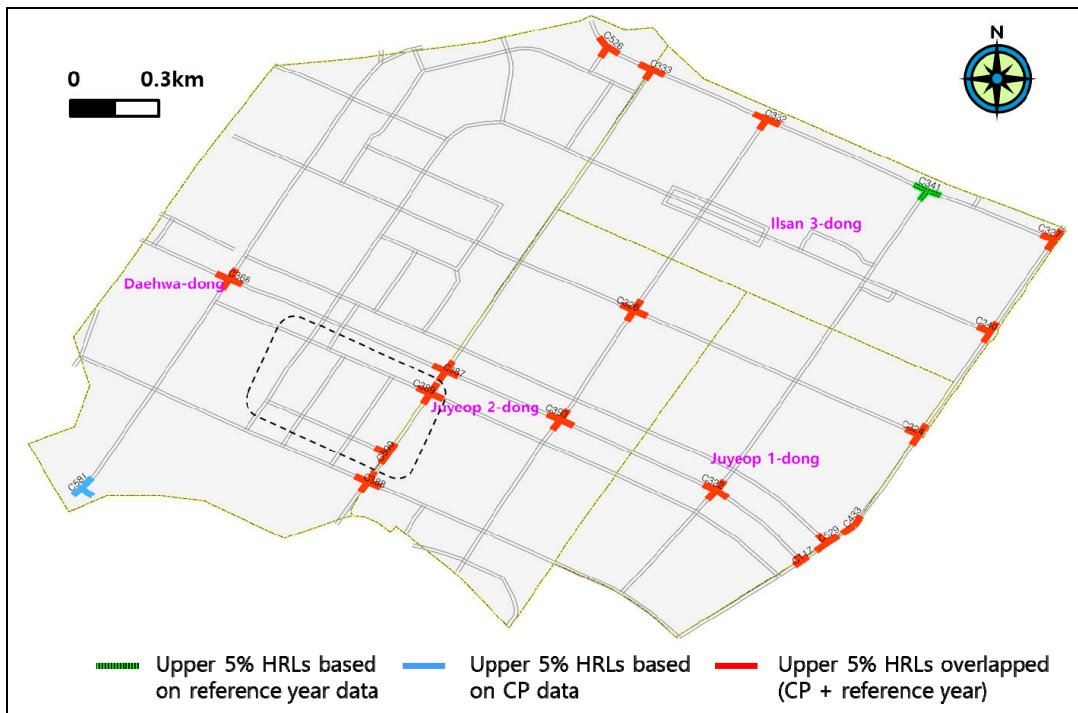
In spite of the short observation period, these results show that the proposed method using resident-reported information performs better than the traditional method to identify the latest hazardous locations. Since the latest crash data from the police could be older than six months, the traditional method may miss some of the latest critical HRLs. However, the proposed method uses resident-reported information right before the assessment period and can efficiently reflect the latest roadway risk information and roadway environment conditions, including the current roadway layout, traffic operation condition, and land use.

The proposed method can also obtain robust results in spite of the short observation period of one week. The police recorded 1028 crashes for the period of 2010–2012, whereas this study collected 444 pieces of latent road risk information in only one week and not three to five years. As a result, the proposed method produced quite accurate results to identify the HRLs. The sample size of the study is not quite small when considering the short period. Therefore, it is concluded that the CP-based latent road risk information can reflect the latest safety environment on the roadway sections, even with a small sample size.

Figure 6 shows one of the identified HRLs in Table 4 on a GIS map. Figure 6a shows the upper 5% of HRLs from the traditional method and their overlapped locations with the reference year. Figure 6b shows the upper 5% of HRLs from the proposed method. In the traditional method and the proposed method, 15 and 17 of the 18 locations are overlapped, respectively. The traditional method did not identify two HRLs, but the proposed method did. The areas near the two associated HRLs are commercial districts, as shown in the dashed box area in Figure 6, meaning that they could have high potential for pedestrian travel. Therefore, it can be interpreted that the traditional method mainly identifies the HRLs on major roadways. However, the proposed method identifies the HRLs on both major and minor roadways, including roadways that are expected to have high pedestrian traffic.



(a) Upper 5% HRLs identified based on the reference year and the past 3 years of data.



(b) Upper 5% HRLs identified based on the reference year and the CP data.

Figure 6. Upper 5% HRLs identified by using previous and proposed approaches.

### 5. Discussion and Conclusions

This study introduced a new framework to observe latent road risk information to identify HRLs based on CP through an online survey broadcasted through social media. A simple comparison with the EB method was conducted to assess the model to identify HRLs in Korea. The assessment results showed that the proposed model performed better than the traditional method. As a result,

we conclude that the latest resident-reported information has high potential to assess current safety conditions on roadways efficiently, in spite of the short observation periods. It could be very beneficial for roadway operators or policy makers to evaluate safety and publish safety-related policy. It could be particularly beneficial when long-term historical data is not available for regions, such as newly constructed areas, and for cities in underdeveloped countries, where long-term crash data are not available. In addition, the proposed approach identifies the HRLs on both major and minor roadways, such as those with high pedestrian traffic, where the traditional approach is likely to miss identification.

Despite the robust results from this research, there are several remaining limitations and further research issues. First of all, it is still questionable whether the proposed method can capture minor or near-miss crashes that are not reported to the police. The reference data for the assessment are also police-reported data, so some of the crashes could still be missing from the records. Thus, reference data containing “true values” are required to assess the performance of different methods for the HRL identification. Second, the proposed method was assessed by comparison with a traditional method. In spite of the temporal stability issue, this study applied two different datasets obtained from two different time periods due to the short research period. Therefore, it is suggested that the same period be used for the comparison. Third, the proposed method can collect minor or near-miss crashes, but it is limited in collecting more comprehensive crash information from the police, such as crash severity, crash-causing factors, road geometry characteristics, vehicle characteristics, and other environmental characteristics. Therefore, the combination of latent risk information from citizens and more comprehensive crash information from the police would lead to more rigorous and reliable analysis results.

We also need further research to determine an appropriate sample size to obtain a reliable and robust result. Additionally, the success of the model would highly depend on residents participating voluntarily. Thus, further study is required to suggest various policies to sustain the model. Although the proposed method is valid, an extensive publicity campaign should be held to promote the method. Such a campaign could include reasons for participating in this program, how to do so, and what the benefits and incentives are. Incentive policies such as discounts on insurance rates for active participants could lead to the creation of a beneficial cycle to form a more efficient and inexpensive program. Moreover, as seen in Table 1, the elderly did not participate in the proposed program since many do not usually use social media or may not even know how to use it. Therefore, another study should be conducted to find ways to obtain latent road risk information evenly from each age group. Since this research was a preliminary study to assess the feasibility of the proposed program, an online survey method broadcasted through social media was used to collect the latent road risk information. Thus, latent road risk information could not be collected in real time. However, the development of a mobile device application that can recognize users’ voices would give various benefits for collecting latent road risk information.

**Author Contributions:** Y.C. designed the proposed program, collected data, and wrote the manuscript. M.W. analyzed the collected data.

**Funding:** This study was funded by the National Research Foundation (NFR) grant funded by the Korea Government (MSIP) (NRF-2017R1A2B4008984).

**Acknowledgments:** This work was supported by the National Research Foundation (NFR) grant funded by the Korea Government (MSIP) (NRF-2017R1A2B4008984).

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Montella, A. A comparative analysis of hotspot identification methods. *Accid. Anal. Prev.* **2010**, *42*, 571–581. [[CrossRef](#)] [[PubMed](#)]
2. Elvik, R. *State-of-the-Art Approaches to Road Accident Black Spot Management and Safety Analysis of Road Networks*; Transportøkonomisk Institutt: Oslo, Norway, 2007.

3. Zegeer, C.V. *Identification of Hazardous Locations on Rural Highways in Kentucky*; No. 392; Kentucky Bureau of Highways: Lexington, Kentucky, 1974; p. 80.
4. Chung, Y.; Chang, I. How accurate is accident data in road safety research? An application of vehicle black box data regarding pedestrian-to-taxi accidents in Korea. *Accid. Anal. Prev.* **2015**, *84*, 1–8. [[CrossRef](#)] [[PubMed](#)]
5. Chung, Y.; Eom, K. *A Proactive Traffic Accident Countermeasure based on ICT and Its Assessment*; The Korea Transport Institute: Goyang, Korea, 2014; p. 17. (In Korean)
6. Heinrich, H.W. *Industrial Accident Prevention: A Scientific Approach*; McGraw-Hill: New York, NY, USA, 1931.
7. Hydén, C. *The Development of a Method for Traffic Safety Evaluation: The Swedish Traffic Conflicts Technique*; Lund University: Lund, Sweden, 1987.
8. Naji, H.; Xue, Q.; Lyu, N.; Wu, C.; Zheng, K. Evaluating the driving risk of near-crash events using a mixed-ordered logit model. *Sustainability* **2018**, *10*, 2868. [[CrossRef](#)]
9. Behnood, A.; Mannering, F.L. The temporal stability of factors affecting driver-injury severities in single-vehicle crashes: Some empirical evidence. *Anal. Methods Accid. Res.* **2015**, *8*, 7–32. [[CrossRef](#)]
10. Behnood, A.; Mannering, F.L. An empirical assessment of the effects of economic recessions on pedestrian-injury crashes using mixed and latent-class models. *Anal. Methods Accid. Res.* **2016**, *12*, 1–17. [[CrossRef](#)]
11. Medury, A.; Grembek, O.; Loukaitou-Sideris, A.; Shafizadeh, K. Investigating the underreporting of pedestrian and bicycle crashes in and around university campuses—A crowdsourcing approach. *Accid. Anal. Prev.* **2017**. [[CrossRef](#)] [[PubMed](#)]
12. Deacon, J.A.; Zegeer, C.V.; Deen, R.C. Identification of hazardous rural highway locations. *Transp. Res. Rec.* **1975**, *543*, 16–33.
13. Norden, M.; Orlansky, J.; Jacobs, H. Application of statistical quality-control techniques to analysis of highway accident data. *Highw. Res. Board Bull.* **1956**, *117*, 17–31.
14. Hauer, E.; Persaud, B.N. Problem of identifying hazardous locations using accident data. *Transp. Res. Rec.* **1984**, *975*, 36–43.
15. Morin, D.A. Application of statistical concepts to accident data. *Highw. Res. Rec.* **1967**, *188*, 72–79.
16. PIARC. *Road Safety Manual: Recommendations from the World Road Association (PIARC)*; Route 2 Market: Harrogate, UK, 2003.
17. Stokes, R.; Mutabazi, M. Rate-quality control method of identifying hazardous road locations. *Transp. Res. Rec.* **1996**, *1542*, 44–48. [[CrossRef](#)]
18. Elvik, R. Evaluations of road accident blackspot treatment: A case of the iron law of evaluation studies? *Accid. Anal. Prev.* **1997**, *29*, 191–199. [[CrossRef](#)]
19. Elvik, R. The predictive validity of empirical Bayes estimates of road safety. *Accid. Anal. Prev.* **2008**, *40*, 1964–1969. [[CrossRef](#)] [[PubMed](#)]
20. Yu, H.; Liu, P.; Chen, J.; Wang, H. Comparative analysis of the spatial analysis methods for hotspot identification. *Accid. Anal. Prev.* **2014**, *66*, 80–88. [[CrossRef](#)] [[PubMed](#)]
21. Lord, D.; Park, P.Y.-J. Investigating the effects of the fixed and varying dispersion parameters of poisson-gamma models on empirical Bayes estimates. *Accid. Anal. Prev.* **2008**, *40*, 1441–1457. [[CrossRef](#)] [[PubMed](#)]
22. Persaud, B.; Lyon, C.; Nguyen, T. Empirical Bayes procedure for ranking sites for safety investigation by potential for safety improvement. *Transp. Res. Rec.* **1999**, *1665*, 7–12. [[CrossRef](#)]
23. Hauer, E. *Observational before-after Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*; Emerald Group Publishing Limited: Oxford, UK, 1997.
24. Hauer, E.; Harwood, D.W.; Council, F.M.; Griffith, M.S. Estimating safety by the empirical Bayes method: A tutorial. *Transp. Res. Rec.* **2002**, *1784*, 126–131. [[CrossRef](#)]
25. Cheng, W.; Washington, S. New criteria for evaluating methods of identifying hot spots. *Transp. Res. Rec.* **2008**, *2083*, 76–85. [[CrossRef](#)]
26. Florin, P.; Wandersman, A. An introduction to citizen participation, voluntary organizations, and community development: Insights for empowerment through research. *Am. J. Commun. Psychol.* **1990**, *18*, 41–54. [[CrossRef](#)]
27. Ratanavaraha, V.; Jomnonkwao, S. Community participation and behavioral changes of helmet use in Thailand. *Transp. Policy* **2013**, *25*, 111–118. [[CrossRef](#)]

28. DeSario, J.; Langton, S. *Citizen Participation in Public Decision Making*; Greenwood Press: Westport, CT, USA, 1987.
29. Webber, M.M. Alternative Styles for Citizen Participation in Transport Planning. *Highw. Res. Rec.* **1971**, *356*, 6–11.
30. Mason, J.B.; Moore, C.T. Development of Guides for Community Acceptance of Highway Location, Development, and Construction. *Highw. Res. Rec.* **1971**, *356*, 43–54.
31. France, E.A. Effects of Citizen Participation in Governmental Decision-Making. *Highw. Res. Rec.* **1971**, *356*, 1–5.
32. Friedmann, J. *Planning in the Public Domain: From Knowledge to Action*; Princeton University Press: Princeton, NJ, USA, 1987.
33. Khisty, C.J. Matching Planning Styles and Roles in Transportation to Conditions of Uncertainty and State Technological Know-How. *Transp. Res. Rec.* **1992**, *1364*, 139–143.
34. Khisty, C.J. Citizen Participation Using a Soft Systems Perspective. *Transp. Res. Rec.* **1992**, *1400*, 53–57.
35. Khisty, C.J. Education and training of transportation engineers and planners vis-à-vis public involvement. *Transp. Res. Rec.* **1996**, *1552*, 171–176. [[CrossRef](#)]
36. Schwartz, M.; Yajima, H.; Bartsch, J.; Brittingham, J. Public involvement in japanese transportation decision making. *Transp. Res. Rec.* **2003**, *1858*, 89–94. [[CrossRef](#)]
37. Ishida, H.; Matsuda, W.; Montalbo, C. People's opinions and road policies: Effects of public involvement activities by the road council in japan. *Transp. Res. Rec.* **1999**, *1685*, 103–112. [[CrossRef](#)]
38. Kikukawa, S.; Yoshioka, M.; Mori, M.; Ishida, H. Drafting japan's new 5-year road improvement and management program: A new approach. *Transp. Res. Rec.* **1999**, *1685*, 144–155. [[CrossRef](#)]
39. Booth, C.; Richardson, T. Placing the public in integrated transport planning. *Transp. Policy* **2001**, *8*, 141–149. [[CrossRef](#)]
40. Bickerstaff, K.; Walker, G. Participatory local governance and transport planning. *Environ. Plan. A* **2001**, *33*, 431–452. [[CrossRef](#)]
41. Bickerstaff, K.; Tolley, R.; Walker, G. Transport planning and participation: The rhetoric and realities of public involvement. *J. Transp. Geogr.* **2002**, *10*, 61–73. [[CrossRef](#)]
42. DETR. *Guidance on Full Local Transport Plans*; DETR, Ed.; Department of the Environment, Transport and the Regions (DETR): London, UK, 2000.
43. Lee, J. *Dissemination of Transport Safety Campaign and Guideline of Citizen Participation I (in Korean)*; Korea Transportation Safety Authority: Ansan, Korea, 1986.
44. Hwang, K. A study on the improvement of citizen participation in the transportation plan and project (in Korean). *J. Urban Stud.* **1996**, 259–272.
45. Kwon, Y.; Yun, S. A Study on Transportation Planning Using Public Involvement (in Korean). In Proceedings of the 49th Korean Society of Transportation Conference, Incheon, Korea, 28–29 October 2005; Korean Society of Transportation, University of Incheon: Incheon, Korea, 2005; pp. 409–415.
46. Kim, H.; Kwon, Y.; Jung, B.; Lee, S. Evaluation of consciousness on public involvement of road project (in Korean). *J. Korean Soc. Transp.* **2006**, *24*, 73–83. [[CrossRef](#)]
47. Grossardt, T.; Bailey, K.; Brumm, J. Structured public involvement: Problems and prospects for improvement. *Transp. Res. Rec.* **2003**, *1858*, 95–102. [[CrossRef](#)]
48. O'Connor, R.; Schwartz, M.; Schaad, J.; Boyd, D. State of the Practice: White Paper on Public Involvement. Available online: [https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwin1qnm8qLeAhUFbFAKHHTudBYYQFjAAegQIBhAC&url=http%3A%2F%2Fwww.des.ucdavis.edu%2Ffaculty%2Fhandy%2FTTP220%2FPI\\_white\\_paper.pdf&usg=AOvVaw0YVELJ3A-wzFKM2X8AuVWM](https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=2ahUKEwin1qnm8qLeAhUFbFAKHHTudBYYQFjAAegQIBhAC&url=http%3A%2F%2Fwww.des.ucdavis.edu%2Ffaculty%2Fhandy%2FTTP220%2FPI_white_paper.pdf&usg=AOvVaw0YVELJ3A-wzFKM2X8AuVWM) (accessed on 25 October 2018).
49. Laituri, M.; Kodrich, K. On line disaster response community: People as sensors of high magnitude disasters using internet gis. *Sensors* **2008**, *8*, 3037–3055. [[CrossRef](#)] [[PubMed](#)]
50. Kelemen, M.; Smith, W. Community and its 'virtual' promises: A critique of cyberlibertarian rhetoric. *Inf. Commun. Soc.* **2001**, *4*, 370–387. [[CrossRef](#)]
51. Ahlqvist, T.; Bäck, A.; Heinonen, S.; Halonen, M. Road-mapping the societal transformation potential of social media. *Foresight* **2010**, *12*, 3–26. [[CrossRef](#)]
52. O'Reilly, T. *What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*; O'Reilly Media: Sebastopol, CA, USA, 2007; p. 23.

53. Frank, M.R.; Mitchell, L.; Dodds, P.S.; Danforth, C.M. Happiness and the patterns of life: A study of geolocated tweets. *Sci. Rep.* **2013**, *3*, 2625. [[CrossRef](#)] [[PubMed](#)]
54. Wood, S.A.; Guerry, A.D.; Silver, J.M.; Lacayo, M. Using social media to quantify nature-based tourism and recreation. *Sci. Rep.* **2013**, *3*, 2976. [[CrossRef](#)] [[PubMed](#)]
55. Hughes, A.L.; Palen, L. Twitter adoption and use in mass convergence and emergency events. *Int. J. Emerg. Manag.* **2009**, *6*, 248–260. [[CrossRef](#)]
56. Sakaki, T.; Okazaki, M.; Matsuo, Y. Earthquake shakes twitter users: Real-time event detection by social sensors. In Proceedings of the 19th International Conference on World Wide Web, Raleigh, NC, USA, 26–30 April 2010; ACM: Raleigh, NC, USA, 2010; pp. 851–860.
57. Vieweg, S.; Hughes, A.L.; Starbird, K.; Palen, L. Microblogging during two natural hazards events: What twitter may contribute to situational awareness. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010; ACM: Atlanta, GA, USA, 2010; pp. 1079–1088.
58. Starbird, K.; Palen, L.; Hughes, A.L.; Vieweg, S. Chatter on the red: What hazards threat reveals about the social life of microblogged information. In Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work, Savannah, GA, USA, 6–10 February 2010; ACM: Savannah, GA, USA, 2010; pp. 241–250.
59. Starbird, K.; Palen, L. Pass it on?: Retweeting in Mass Emergency. In Proceedings of the 7th International Conference on Information Systems on Crisis Response and Management (ISCRAM 2010), Seattle, WA, USA, 2–5 May 2010; International Community on Information Systems for Crisis Response and Management: Seattle, WA, USA, 2010.
60. Spiro, E.; Irvine, C.; DuBois, C.; Butts, C. Waiting for a Retweet: Modeling Waiting Times in Information Propagation. In Proceedings of the 2012 NIPS Workshop of Social Networks and Social Media Conference, Austin, TX, USA, 1 November 2012.
61. Resch, B. People as sensors and collective sensing-contextual observations complementing geo-sensor network measurements. In *Progress in Location-Based Services*; Krisp, J.M., Ed.; Springer: Heidelberg, Germany, 2013; pp. 391–406.
62. Kaplan, A.M.; Haenlein, M. Users of the world, unite! The challenges and opportunities of social media. *Bus. Horiz.* **2010**, *53*, 59–68. [[CrossRef](#)]
63. Näkki, P.; Bäck, A.; Ropponen, T.; Kronqvist, J.; Hintikka, K.A.; Harju, A.; Pöyhtäri, R.; Kola, P. *Social Media for Citizen Participation: Report on the Somus Project*; VTT Publications: Vuorimiehentie, Finland, 2011; p. 131.
64. Ministry of Land, Infrastructure and Transport (MOLIT). *A Master Plan to Improve National Highway Hazardous Locations*; Ministry of Land, Ed.; MOLIT: Sejong, Korea, 2013.
65. Persaud, B.; Lyon, C. Empirical Bayes before–after safety studies: Lessons learned from two decades of experience and future directions. *Accid. Anal. Prev.* **2007**, *39*, 546–555. [[CrossRef](#)] [[PubMed](#)]
66. Powers, M.; Carson, J.L. *Before-after Crash Analysis: A Primer for Using the Empirical Bayes Method*; FHWA/MT-04-002-8117-21; Department of Civil Engineering, Montana State University: Bozeman, MT, USA, 2004; p. 16.
67. Hauer, E. Overdispersion in modelling accidents on road sections and in empirical Bayes estimation. *Accid. Anal. Prev.* **2001**, *33*, 799–808. [[CrossRef](#)]

