Verification of Zone-30-policy effect on accident reduction using propensity score matching method for multiple treatments

Seya, Hajime

Yoshida, Kazuki

Inoue, Satoru

(Citation)
Case Studies on Transport Policy,9(2):693-702

(Issue Date)
2021-06

(Resource Type)
journal article

(Version)
Accepted Manuscript

Verification of Zone-30-Policy Effect on Accident Reduction Using Propensity Score Matching Method for Multiple Treatments

Hajime Seya[1] Graduate School of Kobe University; hseya@people.kobe-u.ac.jp
Kazuki Yoshida Central Japan Railway Company; yossyphoto@gmail.com
Satoru Inoue Ministry of Land, Infrastructure, Transport and Tourism; inoue-s87rm@mlit.go.jp

In this study, the effect of the Zone 30 policy on traffic accident reduction in Hyogo prefecture, Japan, was analyzed. Specifically, with the micro-districts (cho-cho-moku) of Hyogo prefecture as the analysis units, statistical verification was performed for the cross-section data of 2017, i.e., a few years after the introduction of the "30-km/h zones," using the propensity score matching method for multiple treatments. The micro-districts of Hyogo prefecture were classified into three groups: Group 1 with no 30-km/h zones, Group 2 with 30-km/h zones but without physical devices, and Group 3 with 30-km/h zones and physical devices. The primary analysis results were as follows: The number of "Killed" or "Seriously Injured" (KSI) accidents in Group 3 tended to be lower than that in Group 1, whereas there were no significant differences between Groups 1 and 2. Moreover, the total number of traffic accidents in Group 3 was not less than that in Group 1. These results suggest that, at this point, designating an area as a 30-km/h zone, by itself, has negligible effect on reducing traffic accidents; conversely, such areas can effectively reduce the number of KSI accidents when combined with physical devices. In addition, as a verification method, it is empirically demonstrated that utilizing more than two groups is an effective approach for measuring the impact of traffic safety measures.

Key Words: *Zone30; traffic accident; generalized propensity score; multiple treatments*

Acknowledgments

This study was supported by JSPS KAKENHI Grant Numbers 20H02275, 19H02262, and 20H02274. This is part of the results of the University of Tokyo CSIS collaborative research. We would like to express our gratitude to Dr. Pei-fen Kuo (National Cheng Kung University) for providing valuable information on the assessment of traffic safety measures and Dr. Koichi Ushijima (University of Tsukuba) for providing valuable comments. We would also like to thank the Hyogo Prefectural Police Department for providing traffic accident data. The corresponding author states on behalf of all authors that there is no conflict of interest.

[1] Hajime Seya
Departments of Civil Engineering, Graduate School of Engineering, Kobe University
1-1, Rokkodai, Nada, Kobe, Hyogo, 657-8501, Japan,
E-mail: hseya@people.kobe-u.ac.jp; Tel.: +81-78-803-6278

Verification of Zone-30-Policy Effect on Accident Reduction Using Propensity Score Matching Method for Multiple Treatments

Abstract

In this study, the effect of the Zone 30 policy on traffic accident reduction in Hyogo prefecture, Japan, was analyzed. Specifically, with the micro-districts (cho-cho-moku) of Hyogo prefecture as the analysis units, statistical verification was performed for the cross-section data of 2017, i.e., a few years after the introduction of the “30-km/h zones,” using the propensity score matching method for multiple treatments. The micro-districts of Hyogo prefecture were classified into three groups: Group 1 with no 30-km/h zones, Group 2 with 30-km/h zones but without physical devices, and Group 3 with 30-km/h zones and physical devices. The primary analysis results were as follows: The number of “Killed” or “Seriously Injured” (KSI) accidents in Group 3 tended to be lower than that in Group 1, whereas there were no significant differences between Groups 1 and 2. Moreover, the total number of traffic accidents in Group 3 was not less than that in Group 1. These results suggest that, at this point, designating an area as a 30-km/h zone, by itself, has negligible effect on reducing traffic accidents; conversely, such areas can effectively reduce the number of KSI accidents when combined with physical devices. In addition, as a verification method, it is empirically demonstrated that utilizing more than two groups is an effective approach for measuring the impact of traffic safety measures.

Key Words: *Zone 30; traffic accident; generalized propensity score; multiple treatments*

1. Introduction

According to the "Safety on Community Roads with the Zone 30 Policy," issued by the National Police Agency (2018) of Japan, regarding the traffic-related fatalities during a period of 30 days in 2017 in Japan, 15.3% (677 cases) involved cyclists and 36.9% (1636 cases) involved pedestrians. Together, these two types of accidents account for more than half (52.2%) of the traffic accidents. These numbers are considerably higher than those in countries such as France (20.8%), Germany (27.6%), the UK (30.5%), and the USA (18.2%)[1]. Moreover, in Japan, the percentage of car–pedestrian casualties on community roads (less than 5.5-m wide) is approximately 1.8 times higher than those on wider roads. According to the National Police Agency (2019) of Japan, in the ten years before 2010 (the year before the implementation of the Zone 30 policy), the number of traffic accidents on roads wider than 5.5 m decreased by 29.2%, whereas the reduction of that on narrower roads was stagnated at 8.0%.

Under such circumstances, society is becoming more aware of the importance of measures aimed at reducing traffic accidents on community roads. According to the "Survey and Research Report on the Promotion of Zone Policies on Community Roads," issued by the Survey and Research Committee of Promotion of Zone Policies on Community Roads (2011), policies such as School Zone (in 1972), Living Zone (1974), Silver Zone (1988), and Community Zone (1996) were implemented as wide-ranging traffic control policies designed to prevent traffic accidents on community roads in Japan. The Community Zone policy promoted measures that combined traffic rules (e.g., speed limit of 30 km/h and road closures) with road improvements (construction of speed humps, bottlenecks, and sidewalks) in concentrated areas of approximately 25–50 ha, including community areas and elementary school districts. However, because of the wide standard area of the zones and the difficulty in setting them, the policy was not extended to the national level. The Zone 30 policy introduced in 2011 is based on the zone policies in Europe and sets "30-km/h areas" more flexibly regardless of the dimensions, limiting car traffic as much as possible to prioritize pedestrian traffic. It is a safety measure designed to ensure the safe travel of pedestrians on community roads. Particularly, it specifies zones with a speed limit of 30 km/h along with other safety measures that are combined according to necessity and controls the speed within the zones as well as the cars that pass through them.

The speed limit of 30 km/h was set based on reports indicating that the pedestrian fatality rate drastically increases if a car hits a pedestrian at more than 30 km/h (Traffic Bureau of National Police Agency, 2019). In the 30-km/h zones, it is essential to implement a local speed limit of 30 km/h, which may be accompanied by optional measures, such as symbols and road markings on the zone entrance, traffic rules (e.g., road closure, one-way traffic), physical devices, construction and widening of pedestrian walkways, and removal of the center line on roads. In late 2018, there were 3649 30-km/h zones in Japan; however, the country lags behind Europe, where 30-km/h zones with physical devices are common. Among the optional measures, the percentage of physical devices installed is particularly low, at 4.2% (as of late 2016) (Traffic Bureau of National Police Agency, 2018). However, 2015 was a major turning point for district traffic safety measures with the creation of a criterion for physical devices (Hashimoto and Nishiura, 2016).

With the increase in the number of 30-km/h zones, it has become crucial to assess their effect[2]. In the US, traffic safety measures have long been assessed using the empirical Bayes methods (Hauer, 1997; Persaud and Lyon, 2007), which is applied using a before–after comparison framework. In these methods, as the traffic accident risks are calculated based on the number of accidents or fatalities before and after in the control group, this group should ideally have similar geographical and

[1] The statistics of countries other than Japan are of 2016.

[2] In December 2017, the National Police Agency reported the effect of 30-km/h zones as follows: "We compared the number of traffic accidents that occurred during one year before and after the 30-km/h zones (2490 locations across the country until late 2015) were implemented. Consequently, the number of accidents involving both pedestrians and bicycles decreased (respectively, 23.5% and 18.6%)". The problem with this analysis is that it is limited to the before–after comparison of the number of cases within the 30-km/h zones, i.e., as previously mentioned, because the number of traffic accidents has exhibited a downward trend in recent years, even if the number of accidents in the 30-km/h zones decreases, it is not possible to distinguish whether this is due to the overall trend or the Zone 30 Policy.

socioeconomic conditions similar to the treated group where the zone policies have been introduced. However, control groups are often set ad hoc as per the administrative units, raising selection bias concerns.

Nevertheless, if the traffic safety measure implementation date is unknown, it is necessary to rely on cross-section data, and several studies are based on regression models for count data, such as the negative binomial regression model. Other studies have adopted propensity score methods (e.g., Davis, 2000; Sasidharan and Donnell, 2013; Wood et al., 2015; Wood and Donnell, 2017). Park and Saccomanno (2007) found that propensity score matching could reduce selection bias associated with treatment effect estimates. However, the authors focused on the case of a single treatment; thus, additional methods are required for the case of multiple treatments. An attempt was made to fill this gap in this study by using the method developed by Yang et al. (2016).

In the case of Japan, as the introduction date of the 30-km/h zones is often not disclosed (at least not in an easily accessible format), analyses must be performed using cross-section data. However, there are opposing opinions on whether it is possible to obtain reliable results with cross-section data-based analyses resembling the before–after analyses; while some reports have indicated that relatively similar results can be obtained (Wood et al., 2015), cross-section analysis has generated unstable results (values estimated with models) in other reports (Hauer, 2010). Therefore, a consensus has not been reached. Li et al. (2019) conducted a simulation experiment in which they compared the empirical Bayes and propensity score methods for evaluating traffic safety measures and indicated the conditions under which each works best. For the propensity score method, they emphasized the importance of sufficiently large sample sizes and the balance of the covariates of the treated and control groups.

In this study, we focused on zone policies and a traffic calming measure and analyzed their effect when combined with physical devices (e.g., speed hump, bottleneck, slalom, and crank) in Japan, where they are not as developed as they are in Europe. Specifically, this study analyzed the results of the effect of the Zone 30 policy on traffic accident reduction in Hyogo prefecture, Japan, where the city of Kobe with a population of more than 1.5 million is located. Using micro-districts (cho-cho-moku, in Japanese) as units, we statistically verified the cross-section data of Hyogo prefecture in 2017, i.e., a few years after the introduction of the 30-km/h zones, using the propensity score matching method for multiple treatments (Yang et al. 2016). The micro-districts were classified into three groups: Group 1 with no 30-km/h zones, Group 2 with 30-km/h zones but without physical devices, and Group 3 with 30-km/h zones and physical devices. Accordingly, we extracted the combined effect of the 30-km/h zones and physical devices[3], which cannot be easily executed with two standard groups (i.e., "with 30-km/h zones" versus "without 30-km/h zones"). Similar to our investigation, several previous studies have assessed the safety effects of multiple treatments. For instance, Park and Abdel-Aty (2015) assessed the effects of multiple roadside elements on roadway segments (driveway density, poles density, distance to poles, and distance to trees) by estimating crash modification factors using the cross-sectional method. Their results indicated that the safety effects decrease as the density of driveways and roadside poles increases. Additionally, Elvik (2009) conducted an exploratory analysis of models for estimating the combined effects of road safety measures. A study on the impact of a road safety program implemented in Victoria, Australia, between 1990 and 1996 indicated that the impact of safety measures is weakened when these measures are combined with other road safety measures. The present study offers a novel approach to assess the effects of multiple treatments.

The remainder of this paper is structured as follows. Previous related studies are discussed in section 2. The target area of this study, the applied data, and verification method are presented in section 3. Subsequently, the empirical analysis results are presented in section 4. Finally, conclusions and future challenges are presented in section 5.

Although this study is a case analysis focused on Japan, we believe that the results can be used as a reference for other countries, including developing countries planning to expand the implementation of zone policies. Furthermore, as a verification method, it is empirically demonstrated that utilizing more than two groups is an effective approach for measuring the impact of traffic safety

[3] With our data, it was not possible to specify Group 4 without 30-km/h zones but with physical devices. Therefore, Group 4 was included in Group 1. Note that such places, should they exist, lower the impact of the 30-km/h zones.

measures.

2. Literature review

The effect of traffic calming measures has been studied extensively. For example, Ewing et al. (2013) examined the effect of such measures in New York city through before–after comparison and concluded that there was no evidence that traffic calming measures lowered the traffic accident rate. However, overall, numerous studies indicate that traffic calming measures can reduce accidents; according to a meta-analysis by Elvik (2001), zone policies reduced traffic accidents with injuries by 15% on average. In Japan, Hashimoto et al. (2000) indicated that the Community Zone of Mitaka city reduced the number of accidents on main roads by 79%, from 21.5 cases per year to 4.5, suggesting that community zones have a definite effect on the introduced areas.

Many of these previous studies are based on the before–after comparison approach. However, Li and Graham (2016) verified the impact of the installation of 20-mph zones in London using doubly robust estimation, which combines the regression and propensity score models. They used 234 20-mph-zones-setup between 2002 and 2007 as the treated group and areas at least 150 m away from these zones, which could be designated as 20-mph zones, as the control group. The KSI casualties were 24% lesser in the treated group than in the control group.

This study, which aimed to verify the impact of 30-km/h zones using a propensity score method, is similar to that by Li and Graham (2016); however, a significant difference is that it applies the propensity score matching method for multiple treatments proposed by Yang et al. (2016) as the verification method. As the Zone 30 policy is often implemented in combination with physical devices, we selected this method to better quantify the impact of these two elements in combination. In Japan, Inada et al. (2020) measured the effect of 30-km/h zones. To quantify the impact of the Zone 30 policy on the incidence of cyclist and pedestrian injuries through interrupted time-series analysis, they used data on cyclist and pedestrian injuries between 2005 and 2016. The monthly number of deaths and serious injuries per person–time on narrow roads (width < 5.5 m) were compared with those on wide roads (width ≥ 5.5 m). The injury rate ratio was regressed with respect to two predictors: numbers of months after January 2005 and September 2011. The analysis results indicated that by 2016, the cumulative changes in the rate ratio ranged from –0.26 to –0.046, depending on the sex and age, and an estimated number of 1,704 (95% confidence interval = 1,293, 2,198) injuries were prevented. This previous study is a time-series macroanalysis, whereas ours is a microanalysis that uses the position information in the 30 km/h zones and traffic accident data.

3. Data and method

3.1. Data summary

In this study, empirical analysis was performed at the level of micro-districts, classified into the following three groups: Group 1 with no 30-km/h zones (8,598 locations), Group 2 with 30-km/h zones but without physical devices (603 locations), and Group 3 with 30-km/h zones and physical devices (151 locations)[4]. As of December 2019, there were 156 designated zones in Hyogo prefecture. The traffic accident data used in this study are related to the traffic accidents that occurred in Hyogo prefecture between 2012 and 2017[5]. The data include information such as the date and time of the accident, location (longitude and latitude), weather, road format (e.g., crossing, single road), accident type, and parties involved.

The location of the 30-km/h zones is published on the website of the Hyogo Prefectural

[4] Each 30-km/h zone spreads across multiple micro-districts. The existence of devices was checked via Google Street View and local survey. The initial intention was to separate Group 1 into groups with and without physical devices. However, it could not be realized due to difficulty obtaining data. Moreover, the area occupied by 30-km/h zones is different in each town and district, but this aspect was excluded in this study.

[5] Data provided by the Hyogo Prefectural Police Department.

Police Department[6]. We manually mapped the 30-km/h zones using this information and by examining it against the digital road map of Japan on GIS[7]. Fig. 1 shows examples of the mapped 30-km/h zones. At the entrance to some of the 30-km/h zones, there are symbol marks, road signs, and indications (Fig. 2); other cases include physical devices, such as cranks and bottlenecks (Fig. 3). The physical devices in this target area are bottlenecks, which indicate the effect produced by bottlenecks.

[Fig. 1. Example of the 30-km/h zone locations (southern Hyogo prefecture)
(Green: micro-districts with no 30-km/h zones; Yellow: micro-districts with 30-km/h zones;
Red: micro-districts with 30-km/h zones and physical devices)], around here

[Fig. 2. "Zone 30" road sign and 30 km/h speed limit sign
(Picture by Yoshida, captured in May 2017 near Seitoku Elementary School in Nada Ward, Kobe)], around here

[Fig. 3. Physical device in a 30-km/h zone (bottleneck)
(From Google Maps, near Seitoku Elementary School in Nada Ward, Kobe)], around here

A directive entitled "Promotion of Zone 30 Policy" issued by the National Police Agency, which includes the points to be considered when setting zones, states that "30-km/h zones must be set actively, giving priority to areas where the demand from local residents is high. A council formed by local residents, local government, road administrators, and police must ensure that a consensus is reached among the residents." Mimura et al. (2015) used the surrounding land and the ideal performance of community roads for determining the optimal location to set 30-km/h zones; however, in practice, zones were not set based on clear criteria. Moreover, Li and Graham (2016) mentioned that even the 20-mph zones in the UK were not based on clear criteria.

Therefore, in this study, when deciding the covariates that may influence the process of setting 30-km/h zones, we selected factors that, according to several previous studies, affect the number of traffic accidents and casualties: number of previous traffic accidents (A−C), road characteristics (D−G), social-economic variables (H−I), and land use (J) (Table 1). To model the decision-making process involved in the introduction of zone policies, ideally, the covariates need to be considered from a period prior to the introduction of the policies. While we were able to obtain such prior socioeconomic variables, data on road characteristics and land use were from a period when the zone policies were introduced due to the unavailability of prior data. However, as these variables represent the aggregated value of each zone, they may not undergo significant changes with the introduction of measures and are therefore unlikely to affect the analysis results.

Moreover, three accident indices from 2012 were introduced in this study, similar to the study by Li and Graham (2016). Typically, this is because setting zones to dangerous areas is prioritized (Mimura et al., 2015) and omitting this variable may compromise the results, suggesting that the Zone 30 policy increases the number of accidents. In Japan, with the addition of position coordinates to the accident registry in 2012, it is generally difficult to use accident data prior to 2012 for analyses; in addition, acquiring information on the period of the introduction of Zone 30 is challenging[8].

[6] https://www.police.pref.hyogo.lg.jp/traffic/regulation/zone30/index.htm

[7] Data source: "ArcGIS Data Collection Road Network 2015"

[8] In this study, we received information on the introduction date of more than 20% of the total zones (37 locations) in Kobe from the Hyogo Prefectural Police Department and verified whether it was valid to use the number of accident cases in 2012. We determined that all the zones in Kobe, except those in Tarumi Ward, were designated starting in 2013. However, it is necessary to acknowledge that as the introduction date of all the other districts, towns, and villages is unknown, using data from 2012 as the data of previous accidents may underestimate the effect.

Variable “KSI_count_2012” represents the number of KSI accident cases, i.e., the number of accidents involving at least one severely injured person or death. Variable “SI_count_2012” indicates the number of slightly injured (SI) accident cases, involving at least one person with minor injuries. Variable “V-to-P_count_2012” represents the number of vehicle-to-person accidents, which is the total number of traffic accidents excluding single-vehicle, vehicle-to-vehicle, and bicycle-to-vehicle accidents. The descriptive statistics are listed in Table 2.

A study on Zone 30 that focused on the driver was conducted by Dinh and Kubota (2013a), who assessed the speeding behavior of drivers on urban residential streets with 30 km/h speed limits in Japan. They determined that the perceived appropriateness of the speed limit influenced the speeding intention and behavior, but the effect of driver characteristics could not be considered due to inaccessibility of such data. Moreover, as the traffic volume on community roads is not measured in Japan, it was excluded in the analysis. However, as the traffic volume is believed to be correlated to the road characteristics (D−G), social-economic variable (H−I), and land use (J), we assume that it is substituted fairly well.

[Table 1: Covariates in the propensity score method used in this study], around here

[Table 2: Descriptive statistics of the covariates used in this study], around here

3.2. Method

In this study, the propensity score matching method was used to measure the impact of the Zone 30 policy. The standard procedure involves the application of the propensity score matching method with two groups (with Zone 30: 1, without Zone 30: 0); however, as some of the 30-km/h zones include physical devices and others do not, each zone may produce different effects. Therefore, a model with three groups must be considered. As the propensity score matching of two groups is established as an empirical analysis method supported by an excellent book by authors such as Imbens and Rubin (2015), we introduced propensity score matching with three groups in this study.

(1) Propensity score matching method with three groups based on generalized propensity score

The description in this section is based on the study by Yang et al. (2016). Let an arbitrary unit (micro-district) be defined as *i*, and let the treatment of unit *i* be expressed as $W_i \in \mathbb{W} = \{1, \dots, T\}$. Typically, there are only two treatment levels, T = 2; however, in this study, we set T = 3. For each unit, T levels of potential outcomes are considered, which are expressed as $Y_i(w)\ for\ w \in \mathbb{W}$. In addition, the observed outcome for unit *i* is the potential outcome corresponding to the treatment received: $Y_i^{obs} = Y_i(W_i)$. Thus, the generalized propensity score (GPS) can be defined as follows (Imbens, 2000):

$$p(w|\boldsymbol{x}) = Pr(W_i = w \mid \boldsymbol{X}_i = \boldsymbol{x}), \tag{1}$$

where $\boldsymbol{x}$ is a covariate vector. $p(w|\boldsymbol{x})$ should satisfy the overlap assumption $p(w|\boldsymbol{x}) > 0\ \forall\ w, \boldsymbol{x}$.

When the allocation mechanism of the treatment satisfies the condition

$$W_i \perp \left(Y_i(1), \dots, Y_i(T)\right) \mid (p(1|\,\boldsymbol{X}_i), \dots, p(T-1|\boldsymbol{X}_i)), \tag{2}$$

(where $\perp$ is the conditional independence), the mechanism is defined as *strongly unconfounded*. When there are only two treatment levels, conditioning is implemented with a scalar quantity known as the propensity score $p(1|\,\boldsymbol{X}_i)$; however, in the case of multiple treatment levels, it is implemented with a vector defined as $p(1|\, \boldsymbol{X}_i), \dots, p(T-1|\boldsymbol{X}_i)$. Therefore, when T is large, the propensity score does not work as a valid dimension reduction method to balance the covariates. Hence, Imbens (2000) proposed weaker conditions using T indicator variables $D_i(w) \in \{0, 1\}$:

$$D_i(w) = \begin{cases} 1 \; if \; W_i = w, \\ 0 \; otherwise. \end{cases} \tag{3}$$

When the allocation mechanism of the treatment satisfies the condition

$$D_i(w) \perp Y_i(w) \mid p(w|\boldsymbol{X}_i), \tag{4}$$

it is defined as *weakly unconfounded*. Here, the average treatment effect (ATE) is given by

$$\mathbb{E}[Y_i(w') - Y_i(w)] = \mathbb{E}[Y_i(w')] - \mathbb{E}[Y_i(w)] = \\ \mathbb{E}\left[\mathbb{E}\left[Y_i^{obs}|W_i = w', p(w'|\boldsymbol{X}_i)\right]\right] - \mathbb{E}\left[\mathbb{E}\left[Y_i^{obs}|W_i = w, p(w|\boldsymbol{X}_i)\right]\right]. \tag{5}$$

Therefore, under a weak unconfoundedness condition, when the expected value of the potential outcome of a certain treatment is calculated, conditioning is only applied to the propensity score (scalar quantity) corresponding to that treatment. As a result, the dimension can be reduced as in the case of two groups. Yang et al. (2016) proposed the matching method and stratified analysis as the specific implementation methods. However, in this study, we used a matching method demonstrated to be relatively efficient in a simulation experiment conducted by Zhang and Kim (2019).

The GPS matching function for calculating $\mathbb{E}[Y_i(w)]$ is as follows:

$$m_{gps}(w, p) = \arg \min_{j:W_j = w} \left\| p(w|\boldsymbol{X}_j) - p \right\|. \tag{6}$$

where $\|.\|$ denotes a generic metric. Based on this matching function, $Y_i(w)$ is "imputed" by the following equation:

$$\hat{Y}_i(w) = Y^{obs}_{m_{gps}(w, p(w|\boldsymbol{X}_i))}. \tag{7}$$

Hence, sample ATE (i.e., $\hat{\tau}_{gps}(w, w')$) is given by the following equation:

$$\hat{\tau}_{gps}(w, w') = N^{-1} \sum_{i=1}^{N} (Y^{obs}_{m_{gps}(w', p(w'|\boldsymbol{X}_i))} - Y^{obs}_{m_{gps}(w, p(w|\boldsymbol{X}_i))}), \tag{8}$$

where N is the number of units. This equation means that after matching, ATE can be estimated by simply taking the subtraction between imputed values for the groups w and w'.

When the propensity score matching is applied, often, there is a range of covariates that have a small probability of receiving a certain treatment. To improve the overlap in these situations, Crump et al. (2009) proposed the setting of a suitable threshold and the removal of units with excessively high and excessively low propensity scores from the analysis. Thus, the ATE is determined as

$$\tau_{gps}(w, w'|\mathbb{C}) = \mathbb{E}[Y_i(w') - Y_i(w)|\boldsymbol{X}_i \in \mathbb{C}]. \tag{9}$$

The sample analog of eq. (9) can be obtained similar to eq. (8). In this study, the ATE with eqs. (8) and (9) are denoted as "without trimming" and "with trimming," respectively. For the derivation of the variance estimator for inference, refer to the Web-based Supplementary Materials 1 by Yang et al. (2016). We used the R package multilevelMatching (v1.0.0) to estimate eq. (9).

4. Empirical analysis

4.1. Basic aggregation

In the first part of the analysis, the number of traffic accidents was aggregated and subjected to simple before–after comparison (Table 3). The applied categories were as follows: Total (total number of accidents), Vehicle-to-person (vehicle-to-person accidents), Slightly injured (accidents in which at least one person sustained minor injuries), Seriously injured (accidents in which at least one person sustained major injuries but no deaths), and Killed (fatal accidents). The numbers suggest that the effect of the 30-km/h zones is insignificant, but as the covariates distribution (such as those listed in Table 1) are not balanced in each group in terms of the standardized mean difference (%SMD) (Table 4), the differences in the number of cases between the groups cannot be interpreted as the final result. Note that %SMD is the statistic commonly used to examine the covariate distribution balance between the treatment groups. Items with an SMD of 25% or more are considered suspect and may indicate an imbalance for that specific variable (Stuart and Rubin, 2008). Accordingly, several variables were not well balanced.

As the number of "Killed" cases was too small to be analyzed, the empirical analysis below was focused on the number of Total, V-to-P, KSI, and SI. It is important to note that the numbers are in units of "number of accidents" and not people.

[Table 3: Before–after comparison of accident cases in the 30-km/h zones], around here
[Table 4: Result of the %SMD calculated for each group (before matching)], around here

4.2. Model analysis

(1) Analysis results of the propensity score matching method with two groups

Below are the evaluation results of the two standard groups (with 30-km/h zones: 1, without 30-km/h zones: 0) obtained using the propensity score matching method. The propensity score was estimated with a binary logit model (Table 5). The %SMD in Table 5 indicates the balance of the covariate distribution between the treatment groups. All the variables, excluding KSI_count_2012, are significant at a level of 1%. Table 6 depicts the ATE and ATE on the treated (ATT) of the KSI and total variables calculated based on the estimated propensity score. The ATT can be calculated as follows:

$$\hat{\tau}_{ATT} = E[E\{Y_i(1)|W_i = 1, p(W_i|\,\boldsymbol{X}_i)\} - E\{Y_i(0)|W_i = 0, p(W_i|\,\boldsymbol{X}_i)\}], \quad (10)$$

where $(Y_i(1)|W_i = 1)$ is the outcome for treated entity *i*, with the treatment (observed) and $(Y_i(0)|W_i = 0)$ is the outcome for treated entity *i*, without the treatment (unobserved and imputed). These results indicate that when only two groups are used, it is not possible to reject the null hypothesis wherein the effect of Zone 30 is zero.

Table 5: Estimated propensity scores (with a binary logit model)

Table 6: ATE/ATT calculation results with two groups

(2) Analysis result of the propensity score matching method with three groups

We estimated the propensity score for three groups (Group 1 with no 30-km/h zones, Group 2 with 30-km/h zones but without physical devices, and Group 3 with 30-km/h zones and physical devices) by the multinomial logit (MNL) model and calculated the ATE based on the estimated propensity scores.

Table 7 lists the estimation result of the MNL model. Numbers 2 and 3 in the variable names represent the estimated values of Groups 2 and 3, respectively, and Group 1 serves as the baseline. As the distribution of variables "ln (emp_dens+1)" and "ln (pop_dens+1)" was not balanced with the conventionally used SMD 25% level, the square term of these variables was added to the covariates in Table 1 (Model (1)). Table 8 presents the %SMD of each group (after matching). It can be observed that the distributions of "width_under5.5_rate" and "ln (pop_dens+1)" in Group 3 are not balanced with the standard SMD 25% level. Hence, we also estimated using Model (2), to which the interaction terms of "ln (pop_dens+1)" and "width_under5.5_rate" were introduced, and using Model (3) in which these variables were excluded.

In Groups 1 and 2, the distribution of the covariates was well balanced with all the models. For Models (2) and (3), the results of Group 3 are depicted in Table 8. With Model (2), although "width_under5.5_rate" does not meet the SMD 25% level, it is better than Model (1) (here, the square term of "width_under5.5_rate" was not introduced because it worsened the SMD). With Model (3), the SMD 25% level was achieved for all the variables. Fig. 4 shows the histogram diagram of the propensity scores of Groups 1–3 based on Model (3). It indicates that the propensity score itself is well balanced. The results of Models (2) and (3), with which the covariates were balanced, are interpreted below. In addition, the results of Model (1) did not show significant differences.

- Number of previous traffic accidents (A−C)

"V-to-P_count_2012" is significant at 1% in Group 2, suggesting that the selection of the areas in Group 2 was based on the number of previous V-to-P accidents. In Group 3 also, the effect is positive but not significant. "SI_count_2012" is negatively significant at a 1% level. As indicated in Table 3, because the number of V-to-P accidents is considerably lower than that of the vehicle-to-vehicle accidents, several SI accidents are vehicle-to-vehicle accidents. Locations involving frequent vehicle-to-vehicle accidents are situated near highway intersections with high traffic volume, unlike community roads. Therefore, it is possible that locations with numerous SI accidents have a low probability of being designated as 30-km/h zones. Moreover, the results indicate that KSI accidents have negligible effect on the probability of an area being designated as a 30-km/h zone.

- Road characteristics (D−G)

As expected, areas with high road density have high probability of being designated as 30-km/h zones (Groups 2 and 3). In addition, the results indicate that if an area has a high percentage of roads less than 5.5-m wide, it tends to become part of Group 2, and if the percentage of 5.5–13-m-wide roads is high, it tends to become part of Group 3 (13 m or above is the baseline). Moreover, if the intersection density is high, the area has low probability of being designated as a 30-km/h zone in Groups 2 and 3.

- Social-economic variable (H−I)

Overall, the effect of the population density is weak. The square term is significant at a 5% level in Group 3, indicating that physical devices tend to be introduced in areas where the population density is particularly high. The fact that employee density has opposite coefficients in Groups 2 and 3 is interesting. This suggests the importance of setting three groups for analysis.

- Land use (J)

The average slope is negatively significant at a 5% level, which is an intuitive result considering the mountainous terrain of Hyogo prefecture.

Table 9 presents the results of the pairwise ATE of the variables Total, V-to-P, SI, and KSI for each model. The results suggest that the difference in KSI between Groups 3 and 1 and Groups 3 and 2 has a significant negative effect at a 5% level (the former has fewer accidents). As mentioned in section 1, the 30 km/h speed of the Zone 30 policy was set because the pedestrian fatality rate increases considerably if a car hits a pedestrian at more than 30 km/h (Traffic Bureau of National Police Agency, 2019). Hence, it is intuitive that it is effective against KSI. Dinh and Kubota (2013b) attempted to explore driver opinions, attitudes, and behaviors with respect to speeding and driving on urban residential streets with a speed limit of 30 km/h. Through a questionnaire survey, their study confirmed that nearly all drivers had exceeded the speed limit and that they intended to do so in the future, if nothing is changed. The results obtained in our study are consistent with the inferences of the previous studies, i.e., at this point, it is possible that 30-km/h zones are effective only when combined with physical devices; the analysis results of this study suggest the need for further awareness campaigns. However, it is important to note that the combined effect of 30-km/h zones and physical devices cannot be distinguished from the effect of physical devices alone due to data constraints.

The above results were not observed when only two standard groups were used. This may be because when only two groups are used, the effects of Group 2 (ineffective, large share) and Group 3 (effective, small share) are averaged. Although empirical studies with multiple groups in the context of propensity score methods are rare, this study suggests that they can be highly beneficial.

[Table 7: Propensity score estimation results (with multinomial logit model)], around here

[Table 8: Result of %SMD calculated for each group (after matching)], around here

[Fig. 4: Histogram of propensity score (top: Group 1; center: Group 2; bottom: Group 3)], around here

[Table 9: Pairwise ATE results calculated with three groups], around here

(3) Robustness check

For robustness check, we estimated the ordered logit (OL) model instead of the MNL model to estimate propensity scores for multiple treatment. The estimation was implemented using the R package multilevelMatching (v1.0.0), and it does not support trimming for the OL model implemented by Crump et al. (2009). Hence, to isolate the impact of trimming from the difference of model structure (MNL or OL), we used the initially estimated ATE based on the MNL model both for with and without trimming, followed by comparing the latter to the ATE based on the OL model without trimming.

The results are presented in Table 10, which indicates that the impact of the difference in model structure is minor, and our primary result discussed in the previous subsection does not change even if we use the OL model. The with or without trimming has a major impact, and the results suggest that Zone 30 combined with physical devices may reduce V-to-P accidents. However, the results without trimming are less reliable because of lower overlap of covariates.

[Table 10: Robustness check (top: without trimming; bottom: with OL model and without trimming)], around here

5. Conclusion and future challenges

In this study, we verified the effect of 30-km/h zones as traffic calming measures, both alone and combined with physical devices (bottlenecks). Specifically, we conducted a case analysis of Japan using the propensity score matching method for multiple treatments proposed by Yang et al. (2016) with three groups: Group 1 with no 30-km/h zones, Group 2 with 30-km/h zones but without physical devices, and Group 3 with 30-km/h zones and physical devices. The following are the primary analysis results: In Group 3, the number of KSI accidents tended to be lower than that in Group 1, but there were no significant differences between Groups 1 and 2. The total number of traffic accidents was not lower in Group 3 than in Group 1.

The above results suggest that designating an area as a 30-km/h zone will not contribute substantially in the reduction of traffic accidents. Nevertheless, the number of KSI accidents can effectively be reduced when the 30-km/h zones are combined with physical devices. Moreover, from a verification perspective, we empirically demonstrated that using more than two groups is an effective approach to verify the effect of traffic safety measures. This is because when only two groups are used, the effects of Group 2 (ineffective, large share) and Group 3 (effective, small share) are averaged and cannot be properly extracted. The obtained result is consistent with that obtained by Park et al. (2019), which showed that the installation of a physical device (i.e., flashing beacon) at the School Zone speed limit sign was effective in reducing the crash frequency in School Zone areas.

A limitation of this analysis is the possible existence of the positive and negative spatial spillover effect of the 30-km/h zone designation on the surrounding areas (when the stable unit treatment value assumption conditions are not met). However, this is difficult to verify because traffic volume observation on community roads in Japan is not a common practice. It may be possible to measure the spillover effect using probe data or by defining the surrounding areas; however, as the 30-km/h zones correspond only to parts of the micro-districts (the unit of analysis used in this study), most of the positive spillover effect on the surrounding areas (reducing traffic accidents) may be within the respective micro-districts.

Future research will include comparative analysis with the results obtained using methods based on before–after comparison, such as the empirical Bayes methods and difference-in-differences (triple difference) methods, with further data collection, as well as verifying the stability of the results estimated using different analysis units, such as elementary school districts.

The corresponding author states on behalf of all authors that there is no conflict of interest.

H Seya: Literature Research and Review, Manuscript writing and Editing, Data analysis
K Yoshida: Manuscript writing and Editing, Data analysis
S Inoue: Manuscript writing and Editing, Data analysis

References

1. Crump, R.K., Hotz, V.J., Imbens, G.W. and Mitnik, O.A. (2009) Dealing with limited overlap in estimation of average treatment effects, *Biometrika*, 96 (1), 187–199.
2. Davis, G.A. (2000) Accident reduction factors and causal inference in traffic safety studies: A review, *Accident Analysis & Prevention*, 32 (1), 95–109.
3. Dinh, D.D. and Kubota, H. (2013a) Speeding behavior on urban residential streets with a 30 km/h speed limit under the framework of the theory of planned behavior, *Transport Policy*, 29, 199–208.
4. Dinh, D.D. and Kubota, H. (2013b) Drivers' perceptions regarding speeding and driving on urban residential streets with a 30 km/h speed limit, *IATSS Research*, 37 (1), 30–38.
5. Elvik, R. (2001) Area-wide urban traffic calming schemes: A meta-analysis of safety effects, *Accident Analysis & Prevention*, 33 (3), 327–336.
6. Elvik, R. (2009) An exploratory analysis of models for estimating the combined effects of road safety measures, *Accident Analysis & Prevention*, 41 (4), 876–880.
7. Ewing, R., Chen, L. and Chen, C. (2013) Quasi-experimental study of traffic calming measures in New York City, *Transportation Research Record*, 2364 (1), 29–35.
8. Hashimoto, S., and Nishiura, T. (2016) A Study on the approval of local residents in introduction of color pavement in residential street, *Journal of Japan Society of Civil Engineers D3 (Infrastructure Planning and Management)* (in Japanese), 72 (5), I_879–I_888.
9. Hashimoto, S., Sakamoto, K., Takamiya, S., and Kubota, N. (2000) A Study on safety and amenity impact of "Community Zone": Case study in Mitaka city, Tokyo (in Japanese), *Infrastructure Planning Review*, 17, 797–804.
10. Hauer, E. (1997) *Observational Before-after Studies in Road Safety*, Emerald, Bingley.
11. Hauer, E. (2010) Cause, effect and regression in road safety: A case study, *Accident Analysis & Prevention*, 42 (4), 1128–1135.
12. Imbens, G. W. (2000) The role of the propensity score in estimating dose-response functions, *Biometrika*, 87 (3), 706–710.
13. Imbens, G.W. and Rubin, D.B. (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences*, Cambridge University Press, Cambridge.
14. Inada, H., Tomio, J., Nakahara, S., and Ichikawa, M. (2020) Area-wide traffic-calming Zone 30 policy of Japan and incidence of road traffic injuries among cyclists and pedestrians, *American Journal of Public Health*, 110 (2), 237–243.
15. Japan Society of Traffic Engineers (2017), *Revised Edition – Manual of Zone Policies on Community Roads (in Japanese)*, Japan Society of Traffic Engineers, Japan.
16. Li, H. and Graham, D.J. (2016) Quantifying the causal effects of 20 mph zones on road casualties in London via doubly robust estimation, *Accident Analysis & Prevention*, 93, 65–74.
17. Li, H., Graham, D.J., Ding, H. and Ren, G. (2019) Comparison of empirical Bayes and propensity score methods for road safety evaluation: A simulation study, *Accident Analysis & Prevention*, 129, 148–155.
18. Mimura, Y., Hashimoto, S., Shimada, Y., Ando, R., and Yoshiki, S. (2015) Implementation of area speed management considering the land use and the idealized performance of streets – case study of Toyota city (in Japanese), *Journal of Japan Society of Civil Engineers D3 (Infrastructure Planning and Management)*, 71 (5), I_711–I_723.
19. Park, J. and Abdel-Aty, M. (2015) Assessing the safety effects of multiple roadside treatments using parametric and nonparametric approaches, *Accident Analysis & Prevention*, 83, 203–213.
20. Park, J., Abdel-Aty, M. and Lee, J. (2019) School zone safety modeling in countermeasure evaluation and decision, *Transportmetrica A*, 15 (2), 586–601.
21. Park, P.Y.J. and Saccomanno, F.F. (2007) Reducing treatment selection bias for estimating treatment effects using propensity score method, *Journal of Transportation Engineering*, 133 (2), 112–118.
22. Persaud, B. and Lyon, C. (2007) Empirical Bayes before–after safety studies: Lessons learned from two decades of experience and future directions, *Accident Analysis & Prevention*, 39 (3), 546–555.

23. Sasidharan, L. and Donnell, E.T. (2013) Application of propensity scores and potential outcomes to estimate effectiveness of traffic safety countermeasures: Exploratory analysis using intersection lighting data, *Accident Analysis & Prevention*, 50, 539–553.
24. Stuart, E.A. and Rubin, D.B. (2008) Best practices in quasi-experimental design: Matching methods for causal inference. In Osborne, J. *Best Practices in Quantitative Methods* (pp.155–177). Thousand Oaks, CA:Sage.
25. Survey and Research Report on the Promotion of Zone Policies on Community Roads (2011), Survey and Research Committee of Promotion of Zone Policies on Community Roads (in Japanese),
(https://www.npa.go.jp/bureau/traffic/seibi2/kisei/zone30/pdf/houkokusyo.pdf)
26. Traffic Bureau of National Police Department (2019), "Zone 30 Policy" Outline (in Japanese),
(https://www.npa.go.jp/bureau/traffic/seibi2/kisei/zone30/pdf/zone30.pdf)
27. Traffic Bureau of National Police Department (2018), Safety on Community Roads with the "Zone 30 Policy" (in Japanese),
(http://www.mlit.go.jp/road/road/traffic/sesaku/forum/pdf/2-3.pdf#search=%27%E3%82%BE%E3%83%BC%E3%83%B330+%E5%8A%B9%E6%9E%9C%27)
28. Wood, J.S., Donnell, E.T. and Porter, R.J. (2015) Comparison of safety effect estimates obtained from empirical Bayes before–after study, propensity scores-potential outcomes framework, and regression model with cross-sectional data, *Accident Analysis & Prevention*, 75, 144–154.
29. Wood, J. S. and Donnell, E.T. (2017) Causal inference framework for generalizable safety effect estimates, *Accident Analysis & Prevention*, 104, 74–87.
30. Yang, S., Imbens, G.W., Cui, Z., Faries, D.E. and Kadziola, Z. (2016) Propensity score matching and subclassification in observational studies with multi-level treatments, *Biometrics*, 72 (4), 1055–1065.
31. Zhang, D. and Kim, J. (2019) Use of propensity score and disease risk score for multiple treatments with time-to-event outcome: A simulation study, *Journal of Biopharmaceutical Statistics*, 29 (6), 1103–1115.

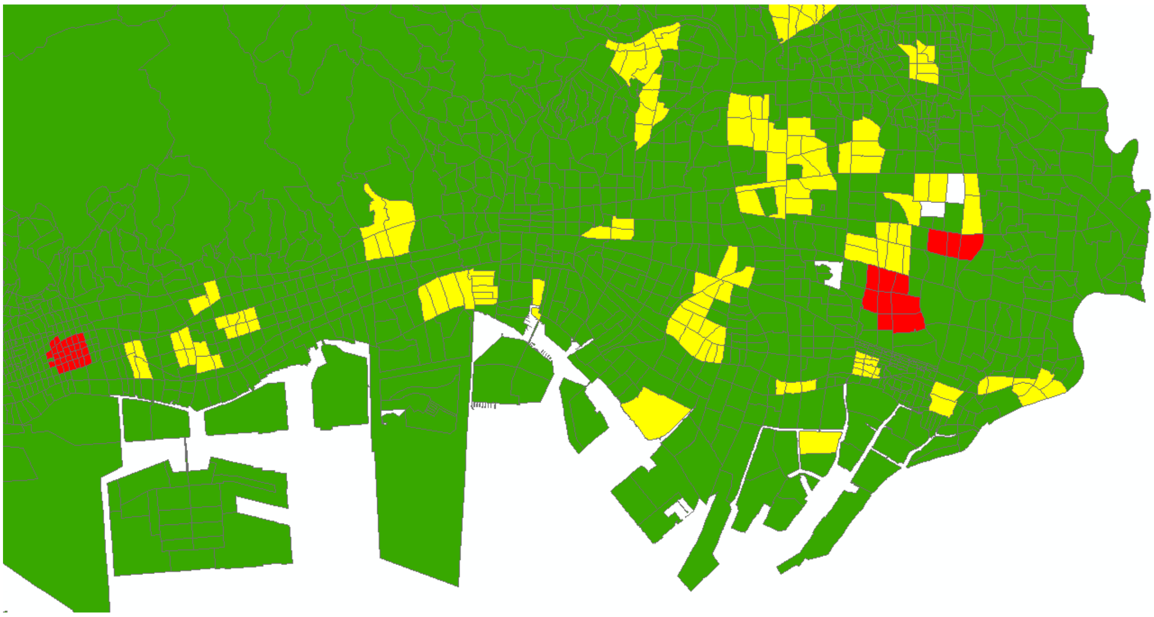


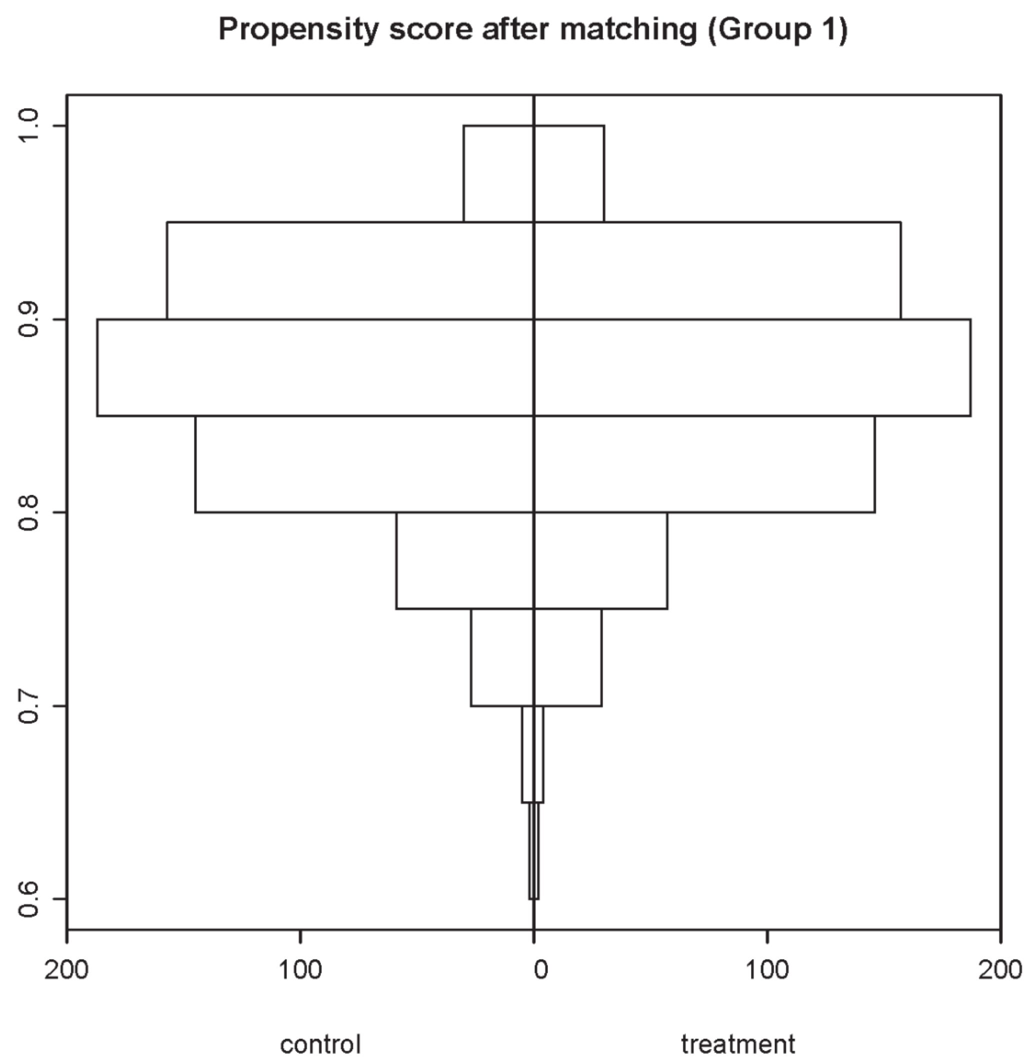
Fig. 1. Example of the 30-km/h zone locations (southern Hyogo prefecture)

(Green: micro-districts with no 30-km/h zones; Yellow: micro-districts with 30-km/h zones; Red: micro-districts with 30-km/h zones and physical devices)

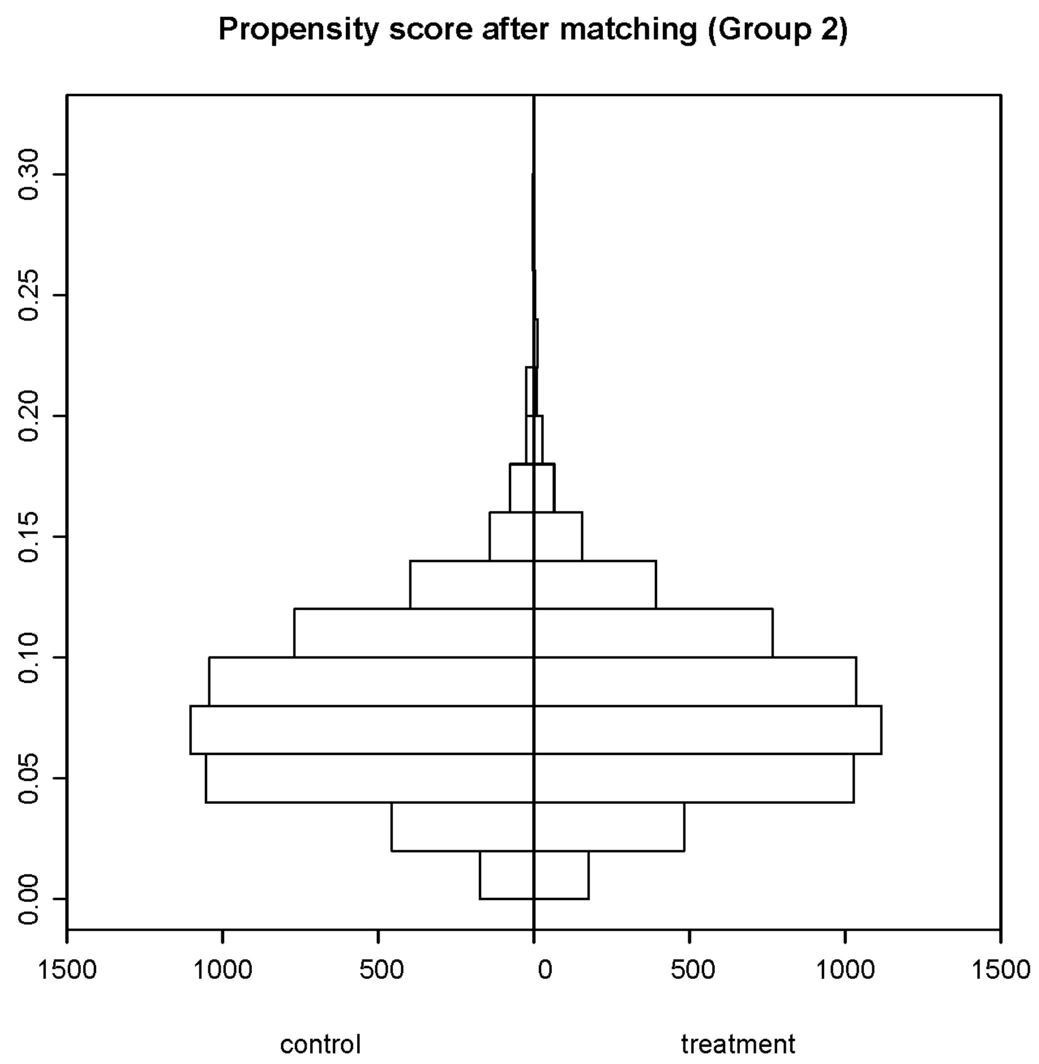
Fig. 2. “Zone 30” road sign and 30 km/h speed limit sign
(Picture by Yoshida, captured in May 2017 near the Seitoku Elementary School in Nada Ward, Kobe)

Fig. 3. Physical device in a 30-km/h zone (bottleneck)

(From Google Maps, near Seitoku Elementary School in Nada Ward, Kobe)



Propensity score after matching (Group 1)
1.0
0.9
0.8
0.7
0.6
200
100
0
100
200
control
treatment


Propensity score after matching (Group 2)
0.00
0.05
0.10
0.15
0.20
0.25
0.30
1500
1000
500
0
500
1000
1500
control
treatment

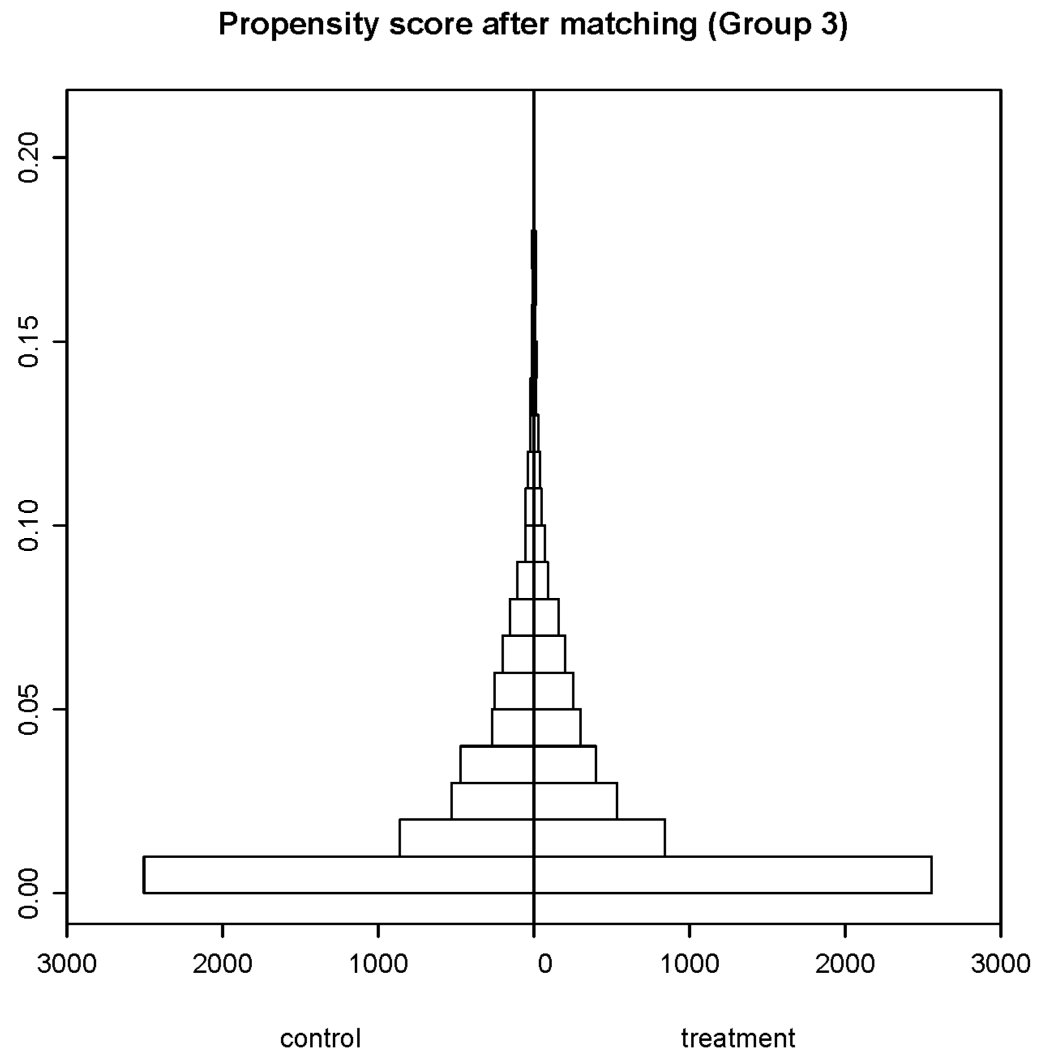


Fig. 4: Histogram of the propensity score (top: Group 1; center: Group 2; bottom: Group 3)
(x-axis: number of cases; y-axis: propensity score estimates)

Table 1: Covariates in the propensity score method used in this study

ID	Variable name	Summary	Unit	Year
A	KSI_count_2012	KSI accident cases in 2012	#	2012
B	SI_count_2012	SI accident cases in 2012	#	2012
C	V-to-P_count_2012	V-to-P accident cases in 2012	#	2012
D	ln (road_dens + ε)	Natural logarithm of the road density + ε in micro-districts	ln [(m / km^2)+ ε]	2015
E	width_under5.5_rate	Percentage of roads in micro-districts less than 5.5 m wide	rate	2015
F	width_5.5_13_rate	Percentage of roads in micro-districts 5.5–13 m wide	rate	2015
G	cross_dens	Intersection density in micro-districts	# / km^2	2015
H	ln (pop_dens + ε)	Natural logarithm of population density + ε in micro-districts	ln [(# / km^2)+ ε]	2010
I	ln (emp_dens + ε)	Natural logarithm of employee density + ε in micro-districts	ln [(# / km^2)+ ε]	2009
J	slope_mean	Average slope (5-m resolution)	degree	2013

Data source: National population census (H), Economic census (I),
ArcGIS Data Collection Road Network (D-G), ArcGIS Data Collection terrain (J).
ε is a small arbitral value to avoid taking the log of zero. We set $\varepsilon = 1$.

Table 2: Descriptive statistics of the covariates used in this study

ID	Variable name	Min.	Median	Mean	Max.
A	KSI_count_2012	0	0	0.3266	9
B	SI_count_2012	0	1	3.179	175
C	V-to-P_count_2012	0	0	0.3234	20
B	ln (road_dens+1)	0	9.840	9.396	17.05
C	width_under5.5_rate	0	0	0.03693	1
D	width_5.5_13_rate	0	0.2367	0.2812	1
E	cross_dens	0	241.0	309.4	5020
F	ln (pop_dens+1)	0	8.396	7.337	15.69
G	ln (emp_dens+1)	0	6.511	5.795	15.51
H	slope_mean	0	0.2626	1.045	14.29

Table 3: Before-after comparison of the accident cases in the 30-km/h zones

	Treatment group		
	1	2	3
Total (2012)	30896	2695	462
Total (2017)	24327	2081	380
Vehicle-to-person (2012)	2758	320	63
Vehicle-to-person (2017)	2555	284	58
Slightly injured (2012)	28039	2422	417
Slightly injured (2017)	22522	1922	354
Seriously Injured (2012)	2700	262	44
Seriously Injured (2017)	1673	148	25
Killed (2012)	166	10	1
Killed (2017)	146	12	1

Treatment groups are Group 1: no 30-km/h zones, Group 2: with 30-km/h zones but without physical devices, and Group 3: with 30-km/h zones + physical devices

Table 4: Result of the %SMD calculated for each group (before matching)

%SMD	Model (3)		
Variables	Group 1 (vs Groups 2 or 3)	Group 2 (vs Groups 1 or 3)	Group 3 (vs Groups 1 or2)
KSI_count_2012	13	17	4.2
SI_count_2012	11	15	7.9
V-to-P_count_2012	22	25	10
ln (road_dens+1)	37	31	62
width_5.5_13_rate	27	15	64
cross_dens	15	9.0	38
ln (emp_dens+1)	51	36	123
ln (emp_dens+1)^2	58	40	136
slope_mean	17	20	4.3

Table 5: Estimated propensity scores (with a binary logit model)

Variables	Estimate	Std. err	z-value	Sig.code	%SMD
(Intercept)	−6.511	0.451	−14.5	***	
KSI_count_2012	0.06405	0.0566	1.13		1.32
SI_count_2012	−0.03093	0.00923	−3.35	***	8.50
V-to-P_count_2012	0.1779	0.0504	3.53	***	6.07
ln (road_dens+1)	0.1541	0.0519	2.97	**	1.67
width_under5.5_rate	1.217	0.417	2.92	**	1.81
width_5.5_13_rate	0.9651	0.162	5.95	***	0.17
cross_dens	−0.0005725	0.000177	−3.24	**	5.92
ln (pop_dens+1)	0.2538	0.0298	8.52	***	3.41
ln (emp_dens+1)	0.08009	0.0202	3.97	***	1.09
slope_mean	−0.1527	0.0305	−5.00	***	3.98
Number of observatories			9712		
McFadden R^2			0.0745		

Significant at ***: 0.1%, **: 1%, *: 5%, x: 10%

Table 6: ATE/ATT calculation results with two groups

Types	Treatment effect	Estimate	Std. err	t-value	p-value
KSI	ATE (caliper 0.25)	0.03084	0.0330	0.935	0.350
KSI	ATT (caliper 0.25)	0.01713	0.0227	0.755	0.450
Total	ATE (caliper 0.25)	0.3116	0.324	0.960	0.337
Total	ATT (caliper 0.25)	−0.1632	0.201	-0.814	0.416

Significant at ***: 0.1%; **: 1%; *: 5%; x: 10% level

“Caliper”, which is the tolerance level of the maximum propensity score distance, is set to 0.25 standard deviation of (the logit of) the propensity score.

As applying logit does not substantially change the estimation results, this study presents the results without applying logit.

Table 7: Estimation result of multinomial logit model

	Model (1)				Model (2)				Model (3)			
Variables	Estimate	Std. err	z-value	Sig.code	Estimate	Std. err	z-value	Sig.code	Estimate	Std. err	z-value	Sig.code
2:(intercept)	-5.398	0.547	-9.86	***	-5.895	0.607	-9.72	***	-5.855	0.548	-10.7	***
3:(intercept)	-13.82	2.14	-6.45	***	-14.15	2.17	-6.51	***	-15.97	2.17	-7.35	***
2:KSI_count_2012	0.09894	0.0602	1.64		0.1012	0.0601	1.68	x	0.09304	0.0595	1.56	
3:KSI_count_2012	-0.04247	0.146	-0.291		-0.03919	0.146	-0.269		-0.05857	0.144	-0.406	
2:SI_count_2012	-0.02006	0.00908	-2.21	*	-0.02031	0.00908	-2.24	*	-0.02293	0.00887	-2.59	**
3:SI_count_2012	-0.1024	0.0303	-3.38	***	-0.1021	0.0304	-3.36	***	-0.1061	0.0298	-3.56	***
3:SI_count_2012	-0.1024	0.0303	-3.38	***	-0.1021	0.0304	-3.36	***	-0.1061	0.0298	-3.56	***
3:V-to-P_count_2012	0.1645	0.123	1.34		0.1678	0.123	1.37		0.1841	0.122	1.51	
2:ln (road_dens+1)	0.1672	0.0558	2.99	**	0.1736	0.0569	3.05	**	0.3360	0.0628	5.35	***
3:ln (road_dens+1)	0.2599	0.137	1.89	x	0.2640	0.136	1.94	x	0.3995	0.147	2.72	**
2:width_under5.5_rate	0.9713	0.453	2.14	*	3.014	0.795	3.79	***	-	-	-	-
3:width_under5.5_rate	0.06404	0.972	0.0659		2.251	1.68	1.34		-	-	-	-
2:width_5.5_13_rate	0.3542	0.190	1.86	x	0.3646	0.191	1.91	x	0.1233	0.185	0.667	
3:width_5.5_13_rate	1.523	0.319	4.77	***	1.541	0.319	4.82	***	1.428	0.309	4.62	***
2:cross_dens	-0.001056	0.000233	-4.54	***	-0.001089	0.000238	-4.57	***	-0.0005244	0.000177	-2.97	**
3:cross_dens	-0.0006077	0.000283	-2.15	*	-0.0006019	0.000283	-2.13	*	-0.0002322	0.000251	-0.927	
2:ln (pop_dens+1)	0.08532	0.105	0.811		0.2007	0.121	1.66	x	-	-	-	-
3:ln (pop_dens+1)	-0.1628	0.154	-1.05		-0.09027	0.170	-0.530		-	-	-	-
2:ln (pop_dens+1)^2	0.01621	0.00810	2.00	*	0.009517	0.00888	1.07		-	-	-	-
3:ln (pop_dens+1)^2	0.03136	0.0129	2.43	*	0.02751	0.0135	2.04	*	-	-	-	-
2:ln (pop_dens+1) * width_under5.5_rate	-	-	-	-	-0.2798	0.0969	-2.89	**	-	-	-	-
3:ln (pop_dens+1) * width_under5.5_rate	-	-	-	-	-0.3104	0.218	-1.43		-	-	-	-
2:ln (emp_dens+1)	-0.2452	0.0502	-4.88	***	-0.2541	0.0502	-5.06	***	-0.1512	0.0492	-3.07	**
3:ln (emp_dens+1)	1.144	0.420	2.72	**	1.132	0.421	2.69	**	1.521	0.424	3.59	***
2:ln (emp_dens+1)^2	0.02501	0.00463	5.41	***	0.02565	0.00462	5.55	***	0.02296	0.00439	5.23	***
3:ln (emp_dens+1)^2	-0.03932	0.0243	-1.62		-0.03854	0.0243	-1.58		-0.05982	0.0244	-2.45	*
2:slope_mean	-0.1917	0.0368	-5.21	***	-0.1907	0.0367	-5.20	***	-0.1484	0.0343	-4.33	***
3:slope_mean	-0.1530	0.0578	-2.65	**	-0.1532	0.0578	-2.65	**	-0.1237	0.0546	-2.27	*
Number of observations						9712						
McFadden R2		0.09456				0.09607				0.07453		

Significant at *** 0.1%, ** 1%, * 5%, x 10%

Table 8: Result of the %SMD calculated for each group (after matching)

%SMD	Model (1)			Model (2)	Model (3)
Variables	Group 1 (vs Groups 2 or 3)	Group 2 (vs Groups 1 or 3)	Group 3 (vs Groups 1 or2)	Group 3 (vs Groups 1 or2)	Group 3 (vs Groups 1 or2)
KSI_count_2012	4.5	1.2	3.4	4.6	14
SI_count_2012	3.5	1.2	2.1	3.9	10
V-to-P_count_2012	1.5	0.6	0.9	1.4	13
ln (road_dens+1)	5.7	9.1	13	20	0.7
width_under5.5_rate	3.1	4.4	45	28	-
width_5.5_13_rate	2.2	3.2	20	14	8.9
cross_dens	7.0	13	8.2	6.4	4.0
ln (pop_dens+1)	4.7	7.5	32	15	-
ln (pop_dens+1)^2	5.9	11	21	1.4	-
ln (pop_dens+1) * width_under5.5_rate	-	-	-	6.6	-
ln (emp_dens+1)	3.1	0.0	13	5.9	3.8
ln (emp_dens+1)^2	3.8	0.6	9.4	3.2	1.2
slope_mean	0.7	7.4	11	4.1	2.8

Table 9: Pairwise ATE results calculated with three groups (top: Model (1); center: Model (2); bottom: Model (3))

Model (1)

Type	Pairwise differences	Mean	Std. err.	z-value	p-value
Total	EY(2) − EY(1)	−0.1542	0.169	−0.914	0.361
	EY(3) − EY(1)	−0.1509	0.858	−0.176	0.860
	EY(3) − EY(2)	0.003266	0.871	0.00375	0.997
V-to-P	EY(2) − EY(1)	0.06132	0.0592	1.04	0.301
	EY(3) − EY(1)	0.006894	0.151	0.0457	0.964
	EY(3) − EY(2)	−0.05443	0.161	−0.338	0.735
Slight injured	EY(2) − EY(1)	−0.1631	0.158	−1.04	0.301
(SI)	EY(3) − EY(1)	−0.03102	0.849	−0.0365	0.971
	EY(3) − EY(2)	0.1321	0.861	0.153	0.878
Killed or Seriously Injured	EY(2) − EY(1)	0.008164	0.0284	0.288	0.774
(KSI)	EY(3) − EY(1)	−0.1196	0.0461	−2.60	0.009
	EY(3) − EY(2)	−0.1277	0.0530	−2.41	0.016

Model (2)

Type	Pairwise differences	Mean	Std. err.	z-value	p-value
Total	EY(2) − EY(1)	−0.09283	0.174	−0.533	0.594
	EY(3) − EY(1)	0.3992	0.591	0.675	0.500
	EY(3) − EY(2)	0.4921	0.611	0.805	0.421
V-to-P	EY(2) − EY(1)	0.05581	0.0451	1.24	0.216
	EY(3) − EY(1)	−0.006566	0.128	−0.0511	0.959
	EY(3) − EY(2)	−0.06237	0.135	−0.461	0.645
Slight injured	EY(2) − EY(1)	−0.1016	0.165	−0.617	0.538
(SI)	EY(3) − EY(1)	0.5253	0.588	0.893	0.372
	EY(3) − EY(2)	0.6268	0.606	1.03	0.301
Killed or Seriously Injured	EY(2) − EY(1)	0.008390	0.0283	0.297	0.767
(KSI)	EY(3) − EY(1)	−0.1257	0.0242	−5.20	0.000
	EY(3) − EY(2)	−0.1341	0.0355	−3.77	0.000

Model (3)

Type	Pairwise differences	Mean	Std. err.	z-value	p-value
Total	EY(2) − EY(1)	−0.1273	0.181	−0.703	0.482
	EY(3) − EY(1)	0.3799	0.509	0.746	0.456
	EY(3) − EY(2)	0.5072	0.535	0.948	0.343
V-to-P	EY(2) − EY(1)	0.03400	0.0428	0.794	0.427
	EY(3) − EY(1)	0.2591	0.191	1.36	0.175
	EY(3) − EY(2)	0.2251	0.195	1.15	0.249
Slight injured	EY(2) − EY(1)	−0.1383	0.173	−0.800	0.424
(SI)	EY(3) − EY(1)	0.5149	0.499	1.03	0.303
	EY(3) − EY(2)	0.6532	0.523	1.25	0.212
Killed or Seriously Injured	EY(2) − EY(1)	0.01081	0.0300	0.360	0.719
(KSI)	EY(3) − EY(1)	−0.1346	0.0348	−3.87	0.000
	EY(3) − EY(2)	−0.1454	0.0447	−3.25	0.001

Table 10: Robustness check (top: without trimming; bottom: estimates from ordered logit model and without trimming)

Model (3)	Without trimming				
Type	Pairwise differences	Mean	Std. err.	z-value	p-value
Total	EY(2) − EY(1)	0.2586	0.374	0.692	0.489
	EY(3) − EY(1)	0.02615	1.06	0.0248	0.980
	EY(3) − EY(2)	− 0.2325	1.12	− 0.208	0.835
V-to-P	EY(2) − EY(1)	0.1065	0.0588	0.809	0.070
	EY(3) − EY(1)	− 0.1415	0.0460	− 3.07	0.002
	EY(3) − EY(2)	− 0.2479	0.0737	− 3.36	0.001
Slight injured	EY(2) − EY(1)	0.2316	0.353	0.656	0.512
(SI)	EY(3) − EY(1)	0.1715	1.05	0.163	0.871
	EY(3) − EY(2)	− 0.06003	1.11	− 0.0541	0.957
Killed or Seriously Injured	EY(2) − EY(1)	0.02698	0.0355	0.759	0.448
(KSI)	EY(3) − EY(1)	− 0.1452	0.0171	− 8.48	0.000
	EY(3) − EY(2)	− 0.1722	0.0385	− 4.47	0.000

Model (3)	Ordered logit, without trimming				
Type	Pairwise differences	Mean	Std. err.	z-value	p-value
Total	EY(2) − EY(1)	0.05694	0.331	0.172	0.863
	EY(3) − EY(1)	− 0.3854	0.917	− 0.420	0.674
	EY(3) − EY(2)	− 0.4423	0.971	− 0.455	0.649
V-to-P	EY(2) − EY(1)	0.06930	0.0528	1.31	0.190
	EY(3) − EY(1)	− 0.1664	0.0327	− 5.08	0.000
	EY(3) − EY(2)	− 0.2357	0.0609	− 3.87	0.000
Slight injured	EY(2) − EY(1)	0.04201	0.311	0.135	0.892
(SI)	EY(3) − EY(1)	− 0.2488	0.916	− 0.272	0.786
	EY(3) − EY(2)	− 0.2908	0.964	− 0.302	0.763
Killed or Seriously Injured	EY(2) − EY(1)	0.01503	0.0342	0.439	0.660
(KSI)	EY(3) − EY(1)	− 0.1364	0.0206	− 6.64	0.000
	EY(3) − EY(2)	− 0.1515	0.0390	− 3.89	0.000