Journal of Japan Association for Earthquake Engineering, Vol. 18, No. 6, 2018

HUMAN VULNERABILITY INDEX FOR EVALUATING TSUNAMI EVACUATION CAPABILITY OF COMMUNITIES

Yozo GOTO[1] **and Tadashi NAKASU**[2]

[1] Member of JAEE, Senior Engineering Adviser, Kaihatsu-Toranomon Consultant Co., Ltd,
Tokyo, Japan, gotoyozo@mti.biglobe.ne.jp
[2] Member of JAEE, Academic Researcher, Chulalongkom University,
Bangkok, Thailand, Tadashi.N@chula.ac.th

ABSTRACT: The rate of fatalities caused by tsunamis vary from community to community depending on geographical and socio-psychological features peculiar to each. If the relationship between fatalities rate and geographical and socio-psychological features can be quantitatively formulated, this can be a concrete means for evaluating a community's vulnerability with regard to evacuation (hereafter, evacuation vulnerability) and developing effective measures that can reduce loss of human life. Therefore, the authors of this paper proposed to apply a Human Vulnerability Index (HVI), defined as fatality rate divided by rate of incidence of washed-out buildings, to evaluate the evacuation vulnerability of municipalities. Using reliable public databases, the authors analyzed the HVIs of twenty municipalities that were heavily damaged by the tsunami of the 2011 Great East Japan Earthquake. Then they applied a multiple-regression analysis using the following four factors as explanatory variables: 1) time allowance for evacuation; 2) preparedness; 3) road serviceability; and 4) warning effect. They thus extracted a reliable formula (R=0.904), which enabled them to quantify the effects of these factors on the HVI. Future tasks are to generalize the formula through application to other tsunami disasters and to establish a numerical evaluation of geographical and socio-psychological features to enable estimation of the tsunami evacuation capability of a municipality and the effect of tsunami countermeasures before a tsunami occurs.

Key Words: Tsunami evacuation, Human Vulnerability Index, The Great East Japan Earthquake, Geographical feature, Socio-psychological feature

1. INTRODUCTION

1.1 Objectives of this study

Many people lost their lives because they could not effectively evacuate from the tsunami of the 2011 Great East Japan Earthquake (GEJE). People's vulnerability to a tsunami, i.e., their lack of ability to escape from areas that are impacted by a tsunami, must essentially be correlated with the geographical and socio-psychological features of each area. Therefore, if it becomes possible to establish a numerical model of such a correlation and then measure the vulnerability of the tsunami-impacted area, it will advance knowledge regarding tsunami-induced human loss and make it possible to evaluate the area's effort to improve its geographical and socio-psychological features. Moreover, generalization of the model to make it applicable to other areas and other tsunami disasters will produce a tool for measuring areal vulnerability to future tsunamis and enable municipalities to prioritize the order of their countermeasures. Hence, the authors have proposed a Human Vulnerability Index (HVI), which is defined as the fatality rate divided by the rate of incidence of buildings damaged, in order to measure the probability of human loss of an area. Furthermore, they have analyzed the correlation between HVI and geographical and socio-psychological features.

The objective of this paper is to verify the hypothesis that HVI is correlated with major geographical and socio-psychological features by establishing a numerical correlation model, and to demonstrate that factors of tsunami evacuation vulnerability can be macro-assessed by this model.

1.2 Literature review

Studies examining features of earthquakes and vulnerability of buildings using the relationship between number of fatalities and number of collapsed buildings have a history of one hundred years. Imamura1) used the ratio of the number of damaged buildings to the number of fatalities to denote the features of an earthquake and presented his findings in a paper for Earthquake Prevention Study Committee Report No.77 published in 1913. Coburn, et al.2) defined the ratio of the number of fatalities to the number of residents in collapsed buildings as a Lethality Ratio, and analyzed the causes of deaths. Ohta, et al.3) improved Kawasumi's equation, $D = 0.01H^{1.3}$ (D: number of fatalities, H: number of collapsed buildings), and proposed a regression equation for D using data from 35 earthquakes from 1872 to 1978. Miyano and Ro4), using data from 1950 and later earthquakes in which most of the human damage was caused by collapsed buildings, proposed a relational expression among hypocentral distance, human damage and building damage. Murakami5) analyzed Lethality Ratios of single-family houses, multifamily houses and non-wooden multifamily houses by applying multiple regression analyses to zone-by-zone numbers of collapsed buildings and fatalities in Ashiya due to the 1995 Great Kobe Earthquake. Matsuda6) defined the number of destroyed houses (collapsed by shaking, burning and being washed away by tsunami) per death as an HD value, and applied this to characterize the features of each earthquake disaster since the Meiji era. Moroi and Takemura7) defined the D/S value as the number of fatalities divided by the number of occupants of collapsed buildings, and applied this to analysis of the features of damage due to the 1995 Kobe Earthquake. Takemura8) analyzed the ratios of the number of destroyed houses (collapsed by shaking, burning, being washed away by tsunami, being buried, and so on) divided by the number of fatalities from major earthquakes since the 1872 Hamada Earthquake, and reported that inland near-field earthquakes caused about one fatality per ten collapsed buildings.

The above studies mainly analyzed the relationship between the shaking of earthquakes and number of fatalities. For the study on tsunami disasters, Kuwasawa, et al.9) surveyed the reactions of people in Owase (one of the tsunami risk-cities in Mie Prefecture) at the time of the 2004 Kii Peninsula Southeast Offshore Earthquake. This earthquake did not cause a destructive tsunami, but the shaking prompted some people to evacuate from the coast. Therefore, Kuwasawa, et al. focused on the consciousness of evacuation and proposed a decision-making model for tsunami evacuation through a multiple regression analysis using five explanatory variables: earthquake intensity, tsunami risk awareness, prejudice of normalization, risk degree of dwellings and preparedness.

After the GEJE, Suzuki and Hayashi10) tried to derive a relationship between human damage and tsunami hazard of the coastal area. They discussed for each local municipality (hereafter, LM) the correlation between fatalities rate, tsunami intensity, geographical features, population exposure, disaster awareness of people and assumed tsunami height. Koyama, et al.11), using 500-meter mesh population data, extracted age and gender distributions of both daytime population and night-time population in the inundated area and in the building washed-out area. They pointed out that the fatality rate rose according to aging of population in LMs and that the missing person rate rose where building-washed-out areas spread over most of the inundated area. Tanishita and Asada12) analyzed the fatalities rates of 59 regions of Minami-sanriku, Miyagi Prefecture, using the tsunami inundation depth, experience from 1960 Chilean Tsunami, viewability of the sea and distance to high land as explanatory variables. They reported that the fatalities rates were high in areas that the 1960 Chilean Tsunami had not inundated and the sea could not be seen.

Goto[13)] proposed a Victim Index, defined as number of fatalities divided by number of damaged buildings, and applied this to six villages in Yamada, Iwate Prefecture. He reported that the rate of tsunami-experienced persons in the village, cognition rate of tsunami warnings, degree of traffic jams, rates of participation in tsunami drills, rate of people who helped others and length of evacuation routes could be related to the Index of each village. Nakasu, et al.[14) 15)] proposed an HVI, defined as fatalities rate divided by collapsed-house rate, as a longitudinal analysis tool. And they applied their HVI to the major municipalities in the Sanriku ria coast of Japan based on their records of damage due to the 1896 Meiji Sanriku Tsunami, the 1933 Showa Sanriku Tsunami and the 2011 Great East Japan Earthquake and Tsunami (GEJET) disaster, and they analyzed the historical shift of vulnerabilities.

On another front, the City Bureau, Ministry of Land, Infrastructure, Transport and Tourism of Japan (MLIT) interviewed 10,603 refugees from the coastal LMs of six prefectures from Aomori to Chiba and asked about their emergency actions[16)]. The City Bureau analyzed the effects of geographical and social features by grouping the LMs into four zones: urban area close to hills; farming-fishing village close to hills; urban area in the plain; and farming-fishing village in the plain. Additionally, effects of traffic jams, risk of human and car mixed evacuations and the effect of steep roads in each LM were analyzed using trip data of evacuees[17)].

1.3 Characteristics and Meanings of this Study

This study is intended to:

(1) redefine Nakasu's HVI by adding the rate of people at home at the time of an earthquake;
(2) measure HVIs of twenty municipalities in Iwate, Miyagi and Fukushima Prefectures numerically using data of the FSC archive (introduced in 2.2.3 and Appendix 1);
(3) verify HVI as an index for measuring tsunami evacuation vulnerability of LMs; and
(4) present a prediction equation for HVI through a multiple regression analysis that uses four characteristic values of geographical and socio-psychological features as explanatory variables.

Expansibility of this study is as follows. If HVI can be generalized to make it adaptable to areas other than the GEJE region, it will be possible to foresee factors causing human damage in risk areas of coming tsunamis, and to numerically evaluate a way to improve evacuation vulnerability.

The uniqueness of this study is the development of a numerical model of actual evacuation vulnerabilities of LMs suffered by the large tsunami of GEJE using site-specific geographical and socio-psychological features. The preceding study by Kuwasawa, et al.[9)] modeled individual decision-making for evacuation using data of an earthquake not accompanied by a destructive tsunami. The study by Suzuki and Hayashi[10)] analyzed damage caused by GEJE with respect to each LM, but only discussed qualitatively the effects of tsunami height, exposed population and assumed scenario tsunami height.

On the other hand, the tsunami evacuation simulation using a multi-agent model can be utilized for disaster education as it can show evacuation action through animation. However, because the effects of geographical and socio-psychological features are included in the simulation, factors affecting evacuation vulnerability cannot be explicitly analyzed.

1.4 Structure of this paper

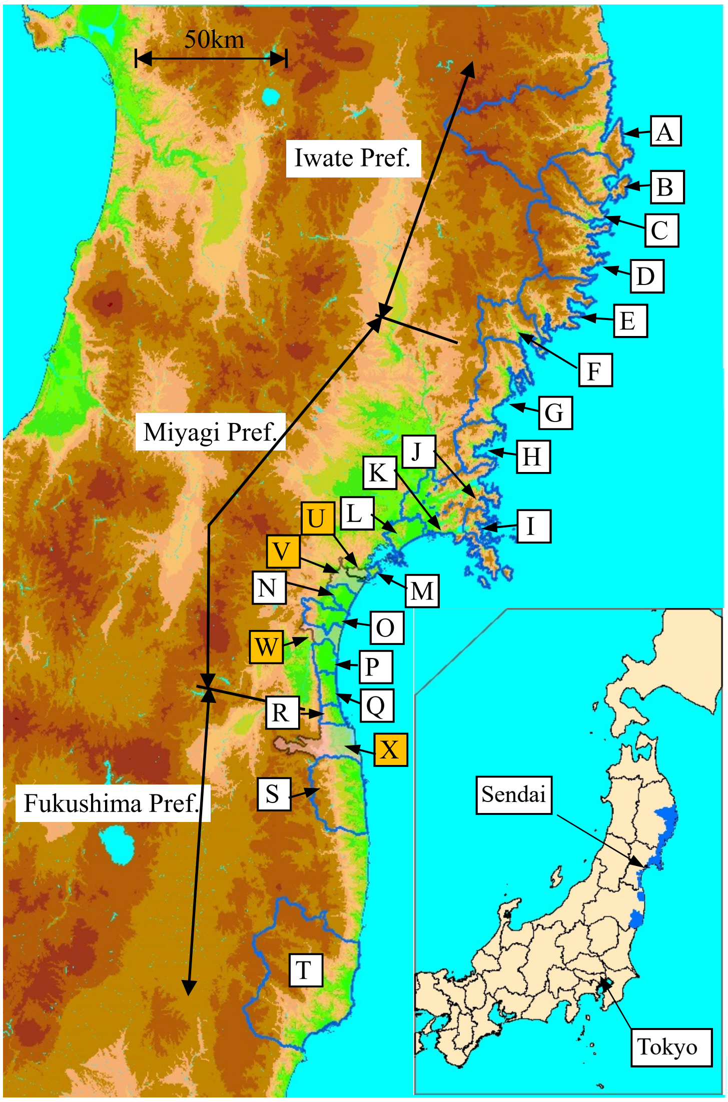
The second chapter describes the analyzed areas, data and their usages. The third chapter describes the formulation of HVI and the calculation procedure, and verifies the hypothesis that HVI, i.e., the quotient of risk and hazard, can measure vulnerability. The fourth chapter verifies the hypothesis that HVI can be defined by geographical and socio-psychological features through multiple regression analysis on the relationship between HVIs and those features. It also validates HVI through sensitivity analysis on the effects of geographical and socio-psychological features to the number of fatalities using the regression equation of HVI. The fifth chapter discusses the availability of some additional explanatory variables and the method of quantification of the geographical and socio-psychological features when HVI is utilized as a prediction tool. The sixth chapter lists conclusions of this study.

The contents of this paper originated from the author's papers for the 2016 JAEE annual conference and the 16th WCEE[18]. However, the formulation of HVI has been revised and discussions added.

2. STUDIED AREAS AND USED DATA

2.1 Analyzed Municipalities

Fig. 1 shows LMs that experienced 78 or more fatalities (including missing persons) due to GEJE, excluding the neighborhood of the Fukushima nuclear power plants. Of these, 20 LMs, from A to T, were selected for this analysis. The LM labeled U in the figure suffered 188 fatalities, but was left out of the analysis because: 1) only a small part of the boundary is facing the sea; 2) almost all of the inundated area was covered by factories and warehouses; 3) 35% of the fatalities were daytime visitors



Drawing from digital map data of Geospatial Information Authority of Japan, and Japan municipality boundary data of ESRI Japan

Fig. 1 Municipalities to be analyzed

from other LMs; and 4) the rate of damaged residential houses was low: 4.7% (the rate for the 20 LMs was 70.1% on average). Additionally, LMs V and W were not included in the target because of the large deviation in age of the limited samples. And LM X was left out because data of people who had undertaken tsunami risk preparedness was estimated unnaturally and was extremely small (these numerical values can be found in footnotes 2) and 3) of Appendix 2).

Another approach would have been to split each LM into areas, such as beach-by-beach or bay-by-bay, so that features of the areas could be analyzed more clearly. However, such splitting would unavoidably reduce the number of available data to less than thirty, and there would have been a loss of significance of the statistical analysis. Therefore, this study did not employ such splitting of LMs.

2.2 Data

2.2.1 Number of fatalities

To obtain the number of fatalities, data from the local governments of Iwate[19)], Miyagi[20)] and Fukushima[21)] Prefectures were utilized. The number of fatalities is defined as the number of dead persons found in the LM area not including indirect deaths, and missing persons counted by the LM. Dead visitors from other LMs were included in the number of fatalities because, in order to analyze evacuation vulnerability of an LM in the daytime, the number of fatalities including visitors should be used. 16 LMs published the number of fatalities of their own citizens. The numbers only differed by 1.9% in total from the number of fatalities defined above and the standard deviation of the difference was 7.8%.

2.2.2 Population and number of houses

The 2010 Census was used to obtain for populations and numbers of houses. As the census does not include the number of houses of LMs I, J, K and R, these numbers were estimated by the method denoted in footnote*3 of Table 2.

2.2.3 Number of damaged houses and inundation depth

Data of the Fukkou-Shien-Chousa archive[22)] (hereafter, the FSC archive) was used to obtain the number of damaged houses and inundation depth data. The FSC archive, outlined in Appendix 1, is a GIS database compiling data from interview surveys conducted by the City Bureau of MLIT.[15)] All the damaged house statistics including inundation depth are in one folder for each LM of the archive. In this study, the number of houses that were washed out by the tsunami was used as the number of damaged houses. For the inundation depth, the average of inundation depths evaluated at the locations of washed-out houses was applied.

There might be some biased counting in the data having fuzzy meaning because many people had to participate to compile the data for different prefectures and there are many LMs in them. Therefore, cumulative addition curves of numbers of washed-out houses were calculated for each LM taking building area as a parameter. The curves were averaged with respect to three prefectures: Iwate (six

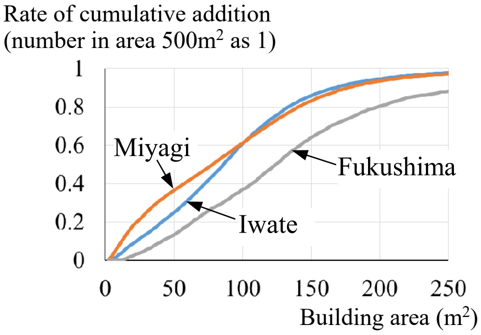


Fig. 2 Cumulative addition curves of washed-out houses vs. building area

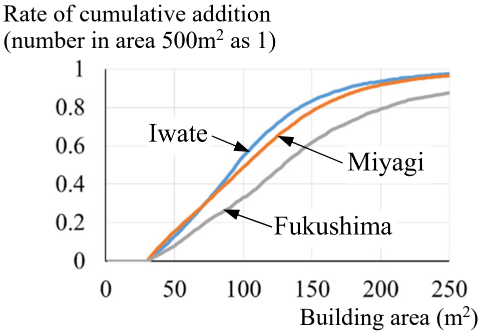


Fig. 3 Cumulative addition curves of washed-out houses vs. building area > 30m²

LMs), Miyagi (eleven LMs) and Fukushima (three LMs). Then, the three cumulative additional curves were drawn, as seen in Fig. 2, taking building area as the horizontal axis and rate of cumulative addition as the vertical axis. Non-negligible discrepancy is seen among the three curves. The counting rules concerning attached small houses, such as barns and garages, seemed to differ among the three prefectures. Therefore, houses of less than 30m^2 in building area were omitted from the cumulative addition and the three curves were re-drawn as shown on Fig. 3. By omitting small houses, the curves of Iwate and Miyagi are closely aligned. Although the curve of Fukushima does not align well, this study decided to use the number of washed-out houses greater than 30m^2 in the building area, taking into account the small number of target LMs in Fukushima.

2.2.4 Sufferers for analysis, rate of people at home and evacuation trip data

Attributes of people who suffered and their evacuation trips were retrieved from the individual evacuation method folder of the FSC archive. The archive compiled individual attributes, evacuation trips and answers to questionnaires covering 10,603 sufferers in the coastal LMs of six prefectures, from Aomori to Chiba. However, in order to focus on the behavior of the persons who might die if they would not evacuate, this study selected the data of people who were at home at the time of the earthquake or returned home before the tsunami arrived, and concurrently whose houses had completely collapsed due to the tsunami. Hereafter, this study denotes these persons as sufferers for analysis.

The rate of people at home was defined as the number of sufferers for analysis divided by the total number of people whose houses were completely destroyed. The values are listed in Table 2 and are consistent with the values of a preceding study[23)]. It should be noted here that the individual attribute data of the FSC archive defined “completely collapsed houses” to include “washed-out houses”.

Evacuation routes and elapsed time for this study were extracted from the trip data of the FSC archive. However, as this was based on interview surveys and not on instrumental observation, it should be recognized that the accuracy is somewhat limited.

2.2.5 Data correction with age

Distribution of age data on individuals in the FSC archive deviated from that of the population. Therefore, age distribution was compared with the census of small segments[24)] which had an average of 137 segments per LM, and correction coefficients with age were defined extracting a group of small segments which covered tsunami inundation area.

Then, the data on individuals in the FSC archive were weighted by the correction coefficient so that the age data distribution came close to that of the census when summing up the data for each LM (for details, see Goto’s past paper[25)]). Table 1 shows the average and standard deviation of the correction coefficients, which were divided into three age ranges: 20-49, 50-69 and over 70. It might have been desirable to divide them into six ranges by adding a sex range. However, such detailed division was not applied because there were too few data.

2.2.6 Discussion on use of survivors' data

While it is more favorable to use data including fatalities for the analysis of evacuation vulnerability, it was impossible to get such data with the same accuracy as for survivors. Therefore, this study used only the data of survivors, as mentioned in 2.2.4.

The ratio of fatalities to sufferers for analysis (defined at 2.2.4), however, needs to be checked. The total number of sufferers for analysis was estimated as 127,000. This was based on the rate of people at home (75%) and the total number of completely destroyed houses in the object region (70,800 houses) with 2.4 people living in one house (2.66 persons per a household in Iwate, Miyagi and Fukushima according to the 2010 Census, with vacant houses estimated to be 10%). The total number of fatalities

Table 1 Average and standard deviation of correction coefficients with age of twenty LMs

Age	20~49	50~69	Over 70
Average	1.209	0.9	1.044
Standard deviation	0.357	0.137	0.185

in the subject region was 16,900. Mikami[26)] reported that 80% of the fatalities were estimated to have been at home or in the process of evacuation. Therefore, the number of fatalities among sufferers for analysis was 13,000, that is, 80% of 16,900. Consequently, the ratio of fatalities to the sufferers for analysis is estimated at around 10%.

If the data concerning fatalities could be added, some aspects of evacuation vulnerability might be emphasized; for example, evacuation roads could be longer, the rate of preparation of emergency carry-out bags would be lower, and so on. Consequently, analysis without data of fatalities might lack sharpness. However, it is not such a simple on-off phenomenon such that evacuation roads longer than a certain length automatically result in evacuation failure. Success or failure of tsunami evacuation depends on many widely dispersed factors. As the data of fatalities were 10% of that of sufferers for analysis, the authors considered the analysis that lacked data of fatalities to not cause a fatal error.

This study might be misunderstood as having developed an evaluation method of the number of fatalities by using survivor data. However, the main subjects of this study are a regression analysis of HVI and its validation. Although evaluation of the number of fatalities is introduced in Chapter 5 of this paper, it is aimed at validating the availability of the regression HVI. Therefore, the validity of using data from survivors in the regression analysis for HVI should be discussed, and this study considered it possible, as mentioned above.

3. FORMULATION OF HVI

3.1 Definition of HVI

This study defined HVI as an index for measuring the evacuation vulnerability of people in an area. It is hypothesized that the number of fatalities (= risk) could be computed by multiplying the exposed population by HVI (= vulnerability) and by inundation depth of a tsunami (= hazard), as described by Eq. (1). For simplification, hazard is expressed by inundation depth, although the flow velocity may affect it.

$$\text{Number of fatalities} \propto \text{Exposed population} \times \text{HVI} \times \text{Inundation depth} \tag{1}$$

HVI is formulated by Eq. (2).

$$\text{HVI} = \frac{\dfrac{\text{Number of fatalities caused by a tsunami}}{\text{Population in area}} \times \dfrac{1}{\text{Rate of people at home}}}{\dfrac{\text{Number of washed-out houses}}{\text{Number of residential houses in area}}} \times 100 \tag{2}$$

Both the number of fatalities and the number of washed-out houses monotonically increase with increased depth of tsunami inundation. However, taking their ratio, HVI becomes independent of inundation depth, as shown by Eq. (2). In addition, the denominator and numerator of Eq. (2) are divided by the number of residential houses and the population in the object area, respectively, enabling HVI to be non-dimensional. Residential houses and population in a common area have to be counted. Therefore, this study used each municipality as the common area, considering data accessibility and versatility. Exceptionally, Ishinomaki was divided into rias coast area and flatland area, and Sendai was divided into administrative wards because of its broadness.

The definition of completely collapsed houses includes washed-out houses. It might have been worth considering applying the number of completely collapsed houses instead of the number of washed-out houses in Eq. (2) in order to fit the definition of "sufferers for analysis" mentioned in subsection 2.2.4. However, as the "number of washed-out houses" of Eq. (2) is used only to evaluate hazard, it is not necessary to use the same definition as in 2.2.4. Therefore, this study used the number of washed-out houses for the reason mentioned in section 3.3.

Multiplication of 100 in Eq. (2) is for improved readability of the HVI value.

3.2 HVI of each LM (Local Municipality)

The HVIs of LMs from A to T of Fig. 1 are listed in Table 2, together with their calculation parameters. The calculated HVIs are scattered from 3 to 33. Chapters 4 and 5 will verify that this scattering means the difference among evacuation vulnerabilities of the LMs.

3.3 Verification of independence of HVI of inundation depth

HVI and inundation depth must be independent of each other in order to verify the hypothesis that HVI represents vulnerability. The right end column of Table 2 lists the average inundation depths evaluated at the locations of washed-out houses in each LM. Fig. 4 is a plot of the HVI of each LM with the averaged inundation depth on the horizontal axis. The dotted line in the figure is the linear regression

Table 2 Parameters and HVI

LM	Fatalities *1	Population*2	Residential houses*3	Rate of people at home*4	Washed-out houses*5	HVI	Inundation depth*6
A	514	59,430	25,010	0.718	1,453	20.7	5.48 (m)
B	752	18,617	7,950	0.729	1,990	22.1	5.40
C	1229	15,276	6,130	0.658	2,989	25.1	7.98
D	1040	39,574	18,420	0.724	2,303	29.0	7.45
E	419	40,737	16,580	0.698	2,397	10.2	7.06
F	1,763	23,300	8,550	0.803	4,210	19.1	11.04
G	1,326	73,489	25,670	0.741	5,817	10.7	7.23
H	812	17,429	5,540	0.752	3,836	8.9	10.71
I	850	10,051	3,450	0.762	2,268	16.9	13.41
J	1,106	23,611	8,105	0.881	3,974	10.8	8.19
K	2,597	137,215	56,765	0.823	4,163	31.4	4.54
L	1,086	42,903	15,450	0.835	2,862	16.4	3.80
M	78	20,416	6,650	0.895	829	3.4	4.54
N	345	132,306	70,640	0.561	1,407	23.3	4.28
O	950	73,134	25,820	0.623	1,888	28.5	4.68
P	270	34,845	11,520	0.819	1,089	10.0	3.37
Q	698	16,704	5,310	0.756	2,061	14.2	5.99
R	99	8,224	3,068	0.827	440	10.1	7.58
S	636	70,878	25,050	0.631	1,083	32.9	5.28
T	330	342,249	147,740	0.733	914	21.3	3.31

*1 Number of dead and missing persons, except related deaths, found in each LM. Iwate, Miyagi and Fukushima Prefectures compiled and published them on the web.

*2 Extracted from 2010 Census data.

*3 Extracted from 2010 Census data; however, the numbers for I and R were not published and for J and K, only sum of J + K was published. Therefore, the numbers for I, R and K were evaluated using the rate of number of residential houses to that of population of neighboring LMs. J was evaluated by subtracting K from the sum.

*4 Rate of people who were in their houses at the time of the earthquake or returned to their houses before the tsunami, and whose houses completely collapsed. Extracted from the FSC archive.

*5 Number of washed-out houses with building areas greater than $30m^2$. Extracted from the FSC archive.

*6 Average of inundation depths evaluated at the locations of washed-out house. Extracted from the FSC archive.

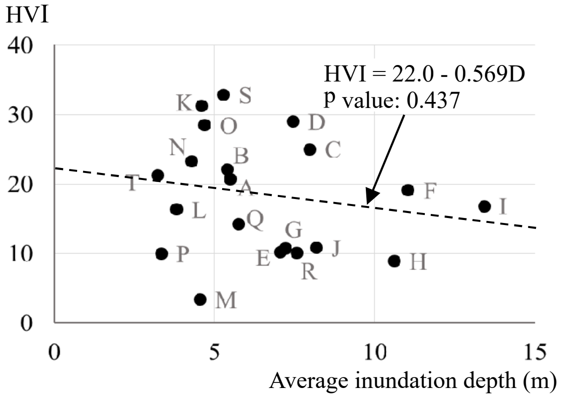


Fig. 4 HVI vs. Average inundation depth

line, which indicates a somewhat decreasing HVI with increased inundation depth. The correlation coefficient of HVI and the average inundation depth was −0.184. However, as the significance level p was 0.437, correlation between HVI and inundation depth was denied from a statistical point of view.

As an alternative, other HVIs were calculated using the number of completely collapsed houses including washed-out houses, instead of the number of washed-out houses only, and analyzed correlation to the average inundation depths evaluated at locations of completely collapsed houses. The correlation coefficient was 0.123 and p was 0.605. Therefore, the independence of HVI would be somewhat improved if the number of completely collapsed houses were used. However, as Takemura[8] pointed out, the definition of completely collapsed house has altered historically and the recent administrative definition seemed to depart from the structural definition to some extent. In addition, considering utilization of overseas data, such a clear definition as washed-out house should be applied.

4. MULTIPLE REGRESSION ANALYSIS ON HVI AND VALIDATION

4.1 Explanatory variables

Many multiple regression analyses on HVIs were executed, taking many different combinations of geographical and socio-psychological features as explanatory variables, as shown in Appendix 2. The following four explanatory variables were extracted as the best combination. The clarity of geographical and socio-psychological meanings of the variables were emphasized in the extraction:

Allowance period A: Tsunami arrival time / Distance to a safe place
Preparedness P: Rate of people who had prepared emergency carry-out bags
Road serviceability R: Rate of car-using evacuees × Car speed
Warning effect We: Tsunami warning height × Cognition rate

4.1.1 Allowance period A

Allowance period A is defined as tsunami arrival time divided by distance to a safe place. Tsunami arrival time means the time for the tsunami to arrive at each LM on the coast after the earthquake and is evaluated as 35 minutes for the rias coast area and 53 - 65 minutes for the flatland area, as shown in Table 3, by referring to previous studies[17), 23)]. Distance to a safe place is the average moving distance of persons in an LM who evacuated their completely collapsed houses, without detour, to high land or inland non-inundation area or vertical evacuation facilities. The individual moving distance was calculated from the individual's evacuation trip data of the FSC archive, using the same method as Goto's preceding study[25)].

When analyzing the average moving distances, it should be noted that in many LMs the number of car-using evacuees was roughly equal to that of walking evacuees. Therefore, the following equivalent evacuation distance, which converted the evacuation distance of a car to that of a pedestrian, was used (Table 3).

$$ed(\mathrm{i}) = \mathrm{W}(\mathrm{i}) \times rw(\mathrm{i}) + \mathrm{D}(\mathrm{i}) \times rc(\mathrm{i}) \times 0.244 \tag{3}$$

where, ed(i): Equivalent evacuation distance of LM i
W(i): Average walking evacuation distances of LM i (average of the distances from house to a safe place)
rw(i): Rate of walking evacuees of LM i
D(i): Average car-driving-evacuation distances of LM i (average of the distances from their house to a safe place)
rc(i): Rate of car-using evacuees of LM i
$0.244 = \Sigma \mathrm{W}(\mathrm{i}) / \Sigma \mathrm{D}(\mathrm{i})$: Conversion factor of car-driving-evacuation distance into walking evacuation distance (average of ratios of all target LMs)

Fig. 5 shows the correlation between HVI and allowance period *A*. The plots are scattered considerably, but HVI tended to increase when the allowance period decreased.

4.1.2 Preparedness P

Preparedness *P* is defined as the rate of sufferers for analysis who had prepared emergency carry-out bags beforehand, and was extracted from personal interview data of the FSC archive (refer Table 3). The

Table 3 HVI and element parameters for explanatory variables

LM	HVI	Tsunami arrival time	Equivalent evacuation distance	Car velocity	Rate of car-using-evacuees	Number of data *ed*, *rc* *1	Tsunami warning height	Cognition rate of warning	Number of data *cr* *2
		t (minute)	*ed* (m)	*v* (km/h)	*rc*		*h*(m)	*cr*	
A	20.7	35	184	8.6	0.504	114	3	0.513	184
B	22.1	35	157	6.4	0.439	35*3	3	0.366	167
C	25.1	35	247	6.4	0.453	90	3	0.368	151
D	29.0	35	175	6.1	0.253	134	3	0.505	209
E	10.2	35	126	8.4	0.586	181	3	0.422	257
F	19.1	35	166	11.1	0.445	145	3	0.566	308
G	10.7	35	247	7.1	0.503	357	6	0.620	490
H	8.9	35	186	7.6	0.551	179	6	0.702	275
I	16.9	35	159	6.2	0.387	94	6	0.589	120
J	10.8	35	135	8.8	0.465	156	6	0.563	236
K	31.4	60	433	6.9	0.487	460	6	0.576	789
L	16.4	60	478	7.7	0.516	68	6	0.430	151
M	3.4	65	166	9.5	0.736	42	6	0.675	57
N	23.3	65	510	13	0.956	31	6	0.501	57
O	28.5	65	501	8.9	0.601	88	6	0.535	155
P	10.0	65	551	18.5	0.703	73	6	0.644	149
Q	14.2	60	425	17.9	0.922	96	6	0.437	134
R	10.1	65	293	13.2	0.721	43	3	0.520	74
S	32.9	65	348	14.1	0.925	97	3	0.301	193
T	21.3	53	190	8.1	0.552	78	3	0.320	110

*1 Number of evacuees who evacuated their houses to a safe place without detour, and whose houses completely collapsed.
*2 Number of sufferers-for-analysis, namely, people who were at their houses at the time of the earthquake or returned to their houses before the tsunami, and whose houses completely collapsed.
*3 A part of the data is missing.

Table 4 Explanatory variables and Regressed HVI

LM	HVI	Allowance period	Preparedness	Road serviceability	Warning effect	Number of data *1	Regression HVI *2
		$A = t / ed$ (minute/m)	P	$R = v \times rc$ (km/h)	$We = h \times cr$ (m)		
A	20.7	0.190	0.498	4.34	1.539	184	14.9
B	22.1	0.222	0.406	2.81	1.098	167	20.7
C	25.1	0.142	0.429	2.90	1.104	151	32.1
D	29.0	0.200	0.342	1.54	1.515	209	30.1
E	10.2	0.278	0.480	4.92	1.266	257	10.3
F	19.1	0.211	0.369	4.94	1.698	308	15.5
G	10.7	0.141	0.403	3.57	3.720	490	18.2
H	8.9	0.188	0.469	4.19	4.212	275	10.3
I	16.9	0.220	0.376	2.40	3.534	120	14.2
J	10.8	0.259	0.361	4.09	3.378	236	10.0
K	31.4	0.139	0.418	3.36	3.456	789	19.1
L	16.4	0.126	0.434	3.97	2.580	151	21.9
M	3.4	0.392	0.410	6.99	4.050	57	4.2
N	23.3	0.127	0.272	12.43	3.006	57	18.5
O	28.5	0.130	0.289	5.35	3.210	155	23.9
P	10.0	0.118	0.412	13.01	3.864	149	12.4
Q	14.2	0.141	0.351	16.51	2.622	134	12.5
R	10.1	0.221	0.414	9.52	1.560	74	10.4
S	32.9	0.187	0.133	13.04	0.903	193	37.0
T	21.3	0.279	0.225	4.47	0.960	110	23.1

*1 Number of evacuees who evacuated directly to a safe place without detour and whose houses completely collapsed.

*2 Regression HVI is discussed in section 4.2.

correlation between HVI and preparedness *P* is shown in Fig. 6. To prepare an emergency carry-out bag beforehand is evidence of consciousness that emergency evacuation might be needed. Considerable correlation was observed as the significance level p was 0.019.

The rates of beforehand-executing of "talking about evacuation method, communication tool, designated place, and so on among family", "checking tsunami hazard map" and "participating in tsunami evacuation drill arranged by the community" were also extracted and were analyzed to determine their correlations with HVI. However, the p values were 0.545, 0.474 and 0.286, respectively, and no significant correlations with HVI were seen.

4.1.3 Road serviceability R

Ideally, road serviceability *R* should be defined by the road traffic capacity, namely, multiplication of the available number of cars and their velocity, in each LM area. However, such data could not be obtained easily. Therefore, data on the actual performance of the tsunami evacuation was used. The better the *R*, the greater the actual number of car-using evacuees and the higher the car velocity are assumed to be. The car velocity *v* and the rate of car-using evacuees *rc* were extracted from the evacuation trip data of the FSC archive, and their multiplication is applied as *R* (Table 3).

Fig. 7 shows the relationship between *R* and HVI. As there are outliers N, Q and S and significance level p is 0.653, no correlation is seen. However, if the outliers are skipped, p becomes 0.039 and correlation is improved, as shown by the red dotted line. As LMs N, Q and S are flatland and low population areas, cars could be driven at their natural velocity. Therefore, the HVIs of these LMs must have been affected strongly by other factors. For example, in LM S, preparedness *P* is 13% (the average of 20 LMs was 37%), and the evacuation rate is 54% (the average was 77%). Hence, many people in LM S could have lacked wariness over tsunamis, and lost their lives without attempting evacuation.

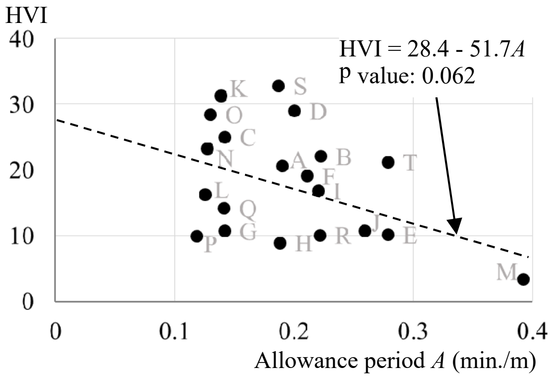


Fig. 5 HVI vs. Allowance period A

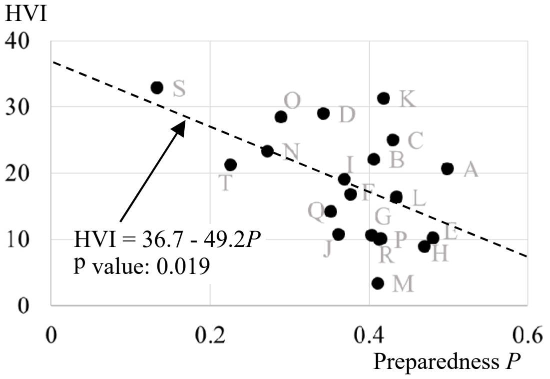


Fig. 6 HVI vs. Preparedness P

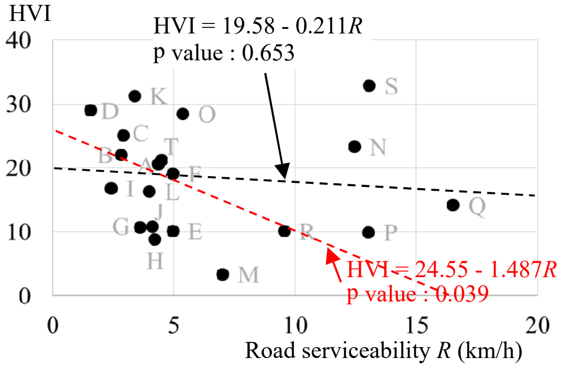


Fig. 7 HVI vs. Road serviceability R

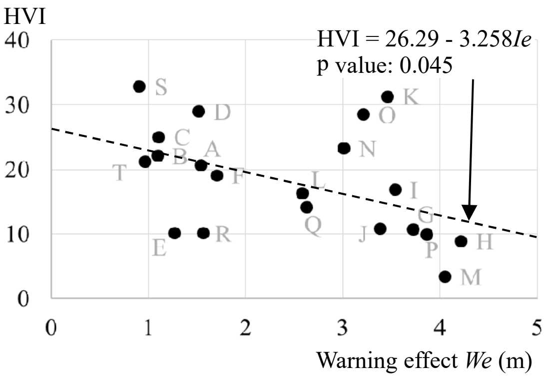


Fig. 8 HVI vs. Warning effect We

4.1.4 Warning effect We

The first announcement of a large-tsunami warning forecast tsunami heights as three meters for the Iwate and Fukushima coasts and six meters for the Miyagi coast. Multiplication of announced tsunami height h and cognition rate of warning cr is defined as Warning effect We. Fig. 8 shows the relationship between We and HVI. Although a certain level of fluctuation is seen, such correlation as significance level p being 0.045 is confirmed.

4.2 Correlation among explanatory variables, and comprehensive influence of geographical features

Correlation coefficients of the four explanatory variables are listed in Table 5. While it is desirable for the correlation coefficients to be low, preparedness P was to some extent related to road serviceability R and warning effect We. Correlation between P and the five element variables v, rc, h, cr and ed were analyzed and the results are listed in Table 6. P indicates a negative correlation to the rate of car-using evacuees rc. Therefore, many of the people who prepared emergency carry-out bags beforehand seemed to have intended to evacuate on foot. P indicates a positive correlation to cognition rate of warning cr. Persons who prepared for evacuation were maintaining a wariness over tsunamis and attention to tsunami warnings. When the sensitivity analysis on the explanatory variables is conducted, these correlations should be considered.

On the other hand, geographical features could have a strong influence. Therefore, a dummy variable that defined the rias coast region (north of Oshika peninsula) as 0 and the flat coast region (southwest of Oshika peninsula) as 1, was introduced and its correlation with the element parameters (Table 3) and the explanatory variables (Table 4) was analyzed.

The results are shown in Table 7. Geographical features are distinctly correlated to tsunami arrival time t and evacuation distance ed, but does not affect allowance period A because the correlations are compensated for through calculation of the ratio between t and ed. Correlation to the rate of car-using evacuees rc is clear. The number of car-using evacuees rc might have increased because the evacuation

Table 5 Correlation factors between explanatory variables

	A	*P*	*R*	*We*
Allowance period : *A*	1	0.040	-0.214	-0.105
Preparedness : *P*	0.040	1	-0.380	0.238
Road Serviceability : *R*	-0.214	-0.380	1	0.058
Warning effect : *We*	-0.105	0.238	0.058	1

Table 6 Correlation factors of preparedness *P* to element variables

	P
Car velocity : *v*	-0.287
Rate of car-using-evacuees : *rc*	-0.415
Height of warned tsunami : *h*	0.087
Cognition rate of warning : *cr*	0.454
Evacuation distance : *ed*	0.250

Table 7 Correlation factors between geographical aspect vs. element and explanatory variables

	t	*ed*	*A*=*t*/*ed*	*P*	*v*	*rc*	*R*=*v*×*rc*	*h*	*cr*	*We*=*h*×*cr*	HVI
Rias or flat	0.981	0.744	−0.143	−0.447	0.563	0.690	0.638	0.302	−0.123	0.138	0.109

distance was long and many roads were easy to drive on in the flat region. However, in conclusion, the geographical features do not correlate with HVI, as p is 0.649 and the correlation coefficient is 0.109. The reasons are inferred as follows. LMs M, R and T are located in the flat coast region, but hills are near the coast, which shortens evacuation distances. While, in LMs K and O, HVIs are pushed up due to traffic jams lowering car velocities.

4.3 Formulation of regression equation and result of multiple regression analysis

A regression equation is formulated as shown Eq. (4). To prevent HVI from being negative, a monomial of exponential terms is applied.

$$\mathrm{HVI} = e^{\alpha} \times A^{\beta} \times P^{\gamma} \times R^{\Delta} \times We^{\varepsilon} \qquad (4)$$

Taking the natural logarithm of both sides of Eq. (4), Eq. (5) is obtained.

$$\ln(\mathrm{HVI}) = \alpha + \beta\times\ln(A) + \gamma\times\ln(P) + \Delta\times\ln(R) + \varepsilon\times\ln(We) \qquad (5)$$

Applying the explanatory variables listed in Table 4, regression coefficients α, β, γ, Δ and ε for Eq. (5) were evaluated by linear multiple regression analysis (IBM SPSS Statistics was used), and the following regression equation for HVI was obtained.

$$\text{Regression HVI} = 2.886 \times A^{-1.117} \times P^{-0.852} \times R^{-0.423} \times We^{-0.441} \qquad (6)$$

The evaluated regression coefficients, their significance probabilities p and standardized coefficients are listed in Table 8. As each p value is less than 0.02, good fit of the analysis is confirmed. The standardized coefficient of β was maximum and that of Δ was second. This means that allowance period *A* is most effective on ln(HVI) and road serviceability *R* is second. VIF, Variance Inflation Factor, is the value that increases if correlation between the explanatory variables becomes higher, and for a VIF of more than 10, the multiple regression analysis becomes unstable because of multicollinearity. As the VIF of each explanatory variable is less than 10, the solution of the multiple regression analysis is confirmed to be stable27).

Table 9 lists evaluation values of the regression equation, with a multiple correlation coefficient R of 0.904 and an adjusted determination coefficient R^2 of 0.768 was achieved. The values of the regression HVIs are calculated from the Eq. (6) and are listed in the right end column of Table 4. Additionally, the plot of the relationship between HVI and the regression HVI is shown in Fig. 9. A trend of similarity is recognized from the plots, even though the standard error is 4.66 and the deviation from the one-to-one line is 12 at maximum.

Table 8 Regression coefficients and evaluation values

	α	β	γ	Δ	ε
Regression coefficient	1.060	−1.117	−0.852	−0.423	−0.441
Significance probability p	0.017	0.000	0.003	0.002	0.004
Standardized coefficient	-----	−0.655	−0.465	−0.477	−0.424
VIF		1.102	1.387	1.252	1.287

Table 9 Evaluation value of regression equation

Correlation coefficient R	0.904
Determination coefficient R^2	0.817
Adjusted R^2	0.768
Standard deviation σ	0.270
Significance probability p	0.000

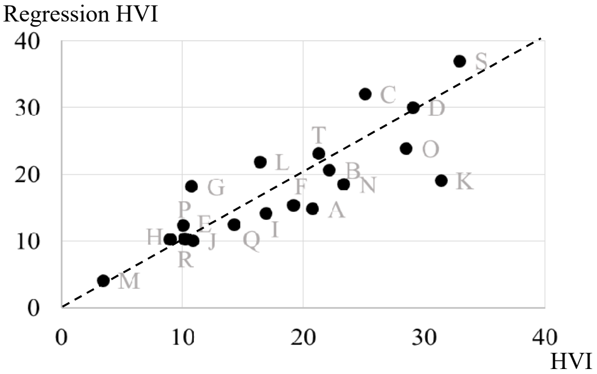


Fig. 9 HVI vs. Regression HVI

4.4 Validation of HVI through sensitivity analysis

In order to estimate the number of fatalities using the regression HVI, Eq. (2) is deformed to Eq. (7). The regression HVI of Table 4 is applied to Eq. (7), and the results are listed in Table 10. The deviation of the estimated number from the actual number is 0.5%.

$$\begin{aligned}\overbrace{\text{Number of fatalities}}^{\textit{Risk}} &= \overbrace{\text{Population} \times \text{Rate of people at home}}^{\textit{Exposed population}} \\ &\times \underbrace{\text{Regression HVI}}_{\textit{Vulnerability}} \times \underbrace{\frac{\text{Number of washed-out houses}}{\text{Number of residential houses}}}_{\textit{Hazard}} \times \frac{1}{100}\end{aligned} \tag{7}$$

In addition, one of the four explanatory variables was multiplied by the rate of increase of standard deviation 1 σ, and regression HVI was calculated keeping the other three variables at their original values. Then, the number of fatalities was evaluated from Eq. (7), and is listed in rows (1) - (4) of Table 11. If the same cost were required to increase each explanatory variable by the rate of 1 σ, increase of allowance period *A* would be the most effective, as already presumed through the comparison of the standardized coefficients of Table 8. If all explanatory variables were increased by the rate of 1 σ, the number of fatalities would decrease by 41%, as shown in row (5) of Table 11.

Table 10 Estimated number of fatalities vs. actual number

	Number of fatalities
Actual number	16,900
Estimated number by Eq. (7)	16,819

Table 11 Sensitivity analysis on explanatory variables

Modified explanatory variable	Number of fatalities
(1) Allowance period A +1σ	12,080
(2) Road serviceability R +1σ	13,540
(3) Preparedness P +1σ	14,090
(4) Warning effect We +1σ	14,220
(5) All explanatory variables +1σ	6,890
(6) P +1σ and warning cognition rate cr +1σ	12,900
(7) P +1σ, warning cognition rate cr +1σ and rate of car-using evacuees rc −1σ	15,130

As shown in Table 7, a positive correlation exists between preparedness P and warning cognition rate cr, which is the element variable of Warning effect We. Therefore, the number of deaths should be calculated when P and cr are increased together. The result is shown in row (6) of Table 11 and proved that the increase of $P + cr$ is as effective as the increase of allowance period A.

Negative correlation existed between preparedness P and the rate of car-using evacuees rc, which was the element variable of road serviceability R. Therefore, the case in which P and cr are increased and rc is decreased by the rate of 1 σ was analyzed. The calculated result is shown in row (7) of Table 11. The decrease of rc has the effect of compensating the increases of P and cr.

For the actual evacuation during the GEJET, P and rc in the LMs studied were in negative correlation. However, if improvements of roads for evacuation, optimization of car usage and sophistication of disaster education were simultaneously implemented, P and rc would not be in negative correlation and their synergistic effect would be expected to develop.

5. DISCUSSION

5.1 Selection of explanatory variables and their combination

The hypotheses that HVI is independent of inundation depth and can be an index for evaluating evacuation vulnerability according to geographical and socio-psychological features are verified through the analyses described in section 3.3 and chapter 4. An inevitable dispersion exists in the data from the interview surveys. Therefore, it can be emphasized that a set of explanatory variables that achieves such high-accuracy regression as correlation coefficient 0.904 was developed. However, this set of explanatory variables does not comply with the sufficient condition, so additional explanatory variables are discussed.

5.1.1 Earthquake intensity

The rates of people of the LMs who immediately thought that a tsunami would come because of the large shaking of GEJE were 49.8% on average, with a standard deviation of 0.173. However, the correlation coefficient between this rate and HVI was as low as −0.188, and multiple regression analysis, including this rate as one of the explanatory variables, did not improve regression accuracy. The JMA seismic intensities of LMs studied were from 5-upper to 6-upper, and the durations of the shaking were similarly long. Therefore, the earthquake intensity that had the effect of prompting people to evacuate did not differ much among the LMs and the differences among the HVIs of the LMs in this study were estimated to be determined by factors other than earthquake intensity.

However, if HVI is applied to a tsunami disaster that is accompanied by an earthquake of different intensity, the effect of earthquake intensity should obviously be added to the multiple regression analysis as one of the explanatory variables.

5.1.2 Aging rate

A high damage rate of aged people is a well-known feature[11), 12)]. Therefore, HVI is presumed to be high in the area of high aging rate. However, the correlation coefficient between the aging rate, which is defined by the ratio of the population over 70 years old to that over 20 years old, and HVI is as low as −0.094. The accuracy of the multiple regression analysis, which was added to the aging rate as one of the explanatory variables, was not improved. Here, the average of the aging rate of 20 LMs was 0.284 and the standard deviation was 0.0256. Therefore, the variation of the aging rates was generally low. Moreover, LM I and LM J, which had relatively higher aging rates, were on the rias coast and had experienced tsunamis frequently. And as reported by the interview survey of Goto and others[22)], the experience seemed to have transmitted from one generation to the next. In such LMs, awareness of tsunamis was shared among the people and there were fewer fatalities. Thus, the HVI might not become high, even if the aging rate is high.

5.1.3 Visibility of sea

Tanishita[12), 28)] reported a tendency in which the number of fatalities increased in areas where there was no direct view of the sea coast. This study could not take into account the visibility of the sea coast because the minimum resolution of the area was the size of each LM. If a finer areal resolution were applied, the visibility of the sea coast would be evaluated by utilizing 3-dimensional GIS and such, and its effect could be studied by adding this as one of the explanatory variables.

5.1.4 Coastal levee

The structural effect of coastal levees is automatically reflected in the hazard term of Eq. (7), because the presence or absence of coastal levees affects the inundation depth and rate of washed-out houses. As for the spiritual effect, people in the area without effective coastal levees might have heightened wariness over tsunamis and evacuate quickly. These effects are reflected in preparedness P.

On another front, taking coastal levees as one of the explanatory variables, it could be possible to analyze the following effects: The presence of coastal levees might reassure people living near-by and induce them to stay in their houses, and they might conceal the tsunami and cause people to delay their evacuation; conversely, it might delay the tsunami inundation. However, these effects of coastal levees could not be analyzed in this study, because no data that covered the location and height of all coast levees of 20 LMs were found.

5.2 Evaluations of key parameters to analyze the evacuation vulnerability of a tsunami-anticipated area using HVI

In order to estimate HVIs for tsunami-anticipated areas and to evaluate evacuation vulnerability, or to analyze the factors that affect the number of fatalities using Eq. (7), the number of washed-out houses and the values of the four explanatory variables must be established in advance.

5.2.1 Number of washed-out houses

The number of washed-out houses can be calculated through the following steps:

(1) Obtain the inundation depth distribution from the results of tsunami inundation simulations that the central or prefecture governments provide to the LMs, and

(2) Calculate the number of washed-out houses using a fragility curve. The fragility curve indicates the relationship between inundation depth and rate of washed-out houses, as proposed by Koshimura, et al.[29)] and others.

However, death by tsunami can happen even if the number of washed-out houses is zero. For these cases, it is necessary to develop another type of explanatory variable that expresses the tsunami hazard. This is a future research issue.

5.2.2 Allowance period A (tsunami arrival time/ evacuation distance)

The tsunami arrival time can be estimated from tsunami simulation conducted by central or prefecture governments. The evacuation distance can be calculated by GIS-aided search of roads that connect starting points with safe places. The former is people's locations when the earthquake strikes, like home, and the latter is the outer area of the assumed inundation area or the vertical evacuation facility.

5.2.3 ***Preparedness P*** *(rate of persons who had prepared emergency carry-out-bags beforehand)*

Data from recent existing surveys can be utilized. In LMs that have not completed such a survey, the data can be collected by questionnaire.

Evaluating the preparation rate of emergency carry-out bags helps to evaluate the level of people's self-directive risk awareness. It should be noted that disaster education such as only urging people to prepare emergency carry-out bags does not directly reduce the number of fatalities.

5.2.4 Road serviceability R *(car velocity x rate of car-using evacuees)*

R is obtained by multiplying car velocity by the rate of car-using evacuees. There are several ways to estimate car velocity: analyzing the road network capacity against car use demand, using a traffic flow simulator to analyze the effect of traffic jams, measuring car velocity at a car use evacuation drill, and so on. The rate of car-using evacuees must be a conceivable rate, not the target rate of a disaster prevention plan.

5.2.5 Warning effect We *(announced tsunami height x cognition rate of the warning)*

Warning effect, *We*, is the equivalent value of the product of announced tsunami height and cognition rate of warning. In the case of GEJET, it was reported that many people in the area where 3 meters was announced as the forecast tsunami height by the first warning received this as a sign of safety and some of them missed the timing of evacuation[30)]. Therefore, 6-meter or 3-meter warning might be a criterion for people to think about evacuation. However, as the effect of the warning to push people to evacuate varies with past tsunami experience and announcement history of forecast height in the area, the effective tsunami height for announcement should be evaluated considering these historical factors.

Cognition rate should not be estimated through analogy of GEJET, because emergency alert emails and other new IT tools to be issued by governmental agencies to disaster forecast areas have been introduced. The effect of such tools should be checked on such occasions as disaster drills.

6. CONCLUSIONS

(1) The Human Vulnerability Index (HVI) introduced by this study was verified to be independent of tsunami height, and to be an index that can express evacuation vulnerability of a studied area being evaluated by geographical and socio-psychological features.

(2) A multiple regression analysis, which set HVI as a target variable and used four explanatory variables, allowance period *A*, preparedness *P*, road serviceability *R* and warning effect *We*, extracted a regression formula that achieved a multi-correlation coefficient of 0.904 and an adjusted coefficient of determination of 0.768. The number of fatalities calculated using the regression HVI deviated from the actual number by only 0.5%.

(3) Sensitivity analysis of the four explanatory variables concerning the number of fatalities indicated allowance period *A* as the most effective and road serviceability *R* as second, if the same variation rate was applied to one of the four variables. Preparedness *P* was closely related to wariness over tsunamis and hence linked with the cognition rate of warnings *cr*, which is the element variable of *We*. Therefore, if these two factors are combined, the effectiveness will be almost the same as that of *A*.

(4) Factors indicated by the sensitivities of these explanatory variables have been qualitatively reported by previous surveys. Nevertheless, by modeling the effects of the variables using HVI, sensitive factors become clear and the efficiency of countermeasures by improving the factors can be numerically evaluated.

(5) Issues for the future are to test HVI for other tsunami disasters and improve the reliability of the regression formula, and to establish a comprehensive evaluation method for explanatory variables in order to apply HVI to other areas subject to tsunami hazard.

ACKNOWLEDGMENT

The City Bureau of Ministry of Land, Infrastructure, Transportation and Tourism of Japan and the Center for Spatial Information Science of the University of Tokyo collected significant data regarding people's evacuation from the Great East Japan Earthquake tsunami and uploaded the data as the Fukkou Shien Chosa archive. The authors would like to express much thanks and respect for their contributions.

Many appreciations to their effective discussions are given to Professor Hitomi Murakami of Yamaguchi University, Professor Masayoshi Tanishita of Chuo University, and the members of the JAEE research committee "Evacuation Research Committee 2012-2016". Associate Professor Maki Koyama of Gifu University is appreciated for her kind advice concerning data exploration. The authors are also grateful to Dr. Raya Muttarak from the International Institute for Applied Systems Analysis (IIASA) for her kind notification of the importance of HVI's potential uses.

APPENDIX 1 Outline of Fukkou Shien Chosa archive (FSC archive)

The FSC archive is a GIS database uploaded by the Center for Spatial Information Science of the University of Tokyo. The original data for the archive is the outcome of the "Survey for Reconstruction of Damaged Cities suffered by the East Japan Great Tsunami, 2011 (Fukkou Shien Chosa in Japanese)" conducted by the City Bureau of Ministry of Land, Infrastructure, Transportation and Tourism of Japan. In line with the contents of the survey report16), the FSC archive web page21), and the paper by Sekimoto, et al.31), the outline of the archive is described below.

1. The database lists individual persons and business offices in 62 local municipalities (hereafter LMs), on the coast from Aomori to Chiba Prefectures. The numbers of samples are 10,603 individuals and 985 business offices. In the survey of the individuals, investigators visited shelters, temporary houses and partially damaged houses, and interviewed people who were affected by the tsunami. The survey term was from September to December of 2011.

2. The sample rate of the individuals was 1.5% - 3% of over 20 age population in the inundation area of LMs. The minimum number of samples in an LM was 20, and in reverse, if number exceeded 500, the sample rate was gradually decreased and the number of samples was limited to around 1,500 at most. In addition, the affected area was divided into two zones, houses completely collapsed and houses partially damaged and inundated, and the number of samples was allocated to be proportional to the population of each area.

For the people in the area where houses had completely collapsed, those in temporary houses were interviewed, and for the people in the area where houses were partially damaged and inundated, those in their own houses were interviewed.

The sex and age distribution of samples were targeted to be similar with those of the population in each LM. However, the actual distribution of samples somewhat deviated from the targets in the LMs studied. There were fewer 20-39 year old males and females and 40-59 year old males. Conversely, there were more over 60 year old males and females, and 40-59 year old females.

3. The archive consists of open data and semi-open data, and governmental or research users can access the latter through a registration procedure. However, for downloading the data, the users are required to promise to take measures to protect personal information and privacy in future publication of their study results.

4. The archive contains a GIS data definition document and each LM's past reconstruction plan as well as each LM's database of many kinds of damage of the 2011 Great East Japan Earthquake. Namely, inundation area, inundation depth, inundation trace, damage overview, building damage, public infrastructure damage (river, coast, steep slope, erosion control facility, windbreak storm surge forest, road, port, sewer, park, and green space), lifeline damage (water and gas), public service damage (bus,

hospital and welfare), cultural asset and educational facility damage, and sufferer and their evacuation manner data are compiled in the database.

5. Outline of data used in this study:

(1) Inundation area data: Polygon data.

(2) Damaged building data: Polygon data of all buildings in the inundation area with attributes such as floor area, structure, usage, classification of damage, year build*, residential house or non-residential house*, adequacy for the use as evacuation points, and inundation depth at building location. (* means data of some MLs are missing).

(3) Evacuation trip data of individual sufferers: Polyline data of individual evacuation trips with attributes such as staying time, start time, arrival time, movement method, purpose of trip, trigger of evacuation start, and tsunami visibility.

(4) Refuge place data of individual sufferers: Point data of refuge place with attributes such as name and type.

(5) Sufferer data of each administrative area: Polygon data of administrative areas in each municipality with attributes such as population before disaster, number of households before disaster, number of deaths, number of missing, number of deceased visitors, and number of decreased households. However, it should be noted that these data were collected before the end of June of 2011 and a considerable number of items are missing.

(6) Interview data of individual sufferers: At the top of the evacuation action sub-folder, a table of interviews of individual sufferers is uploaded. The contents are:

(a) whereabouts of the interviewee at the time of the earthquake,
(b) number of stories of the building where the interviewee was in at the time of the earthquake,
(c) whereabouts of the interviewee's family at the time of the earthquake,

(d) anticipated tsunami coming just after the earthquake or not,
(e) damage of the place where the interviewee was located,
(f) actions the interviewee took after the earthquake,
(g) heard the large tsunami warning or not, heard the height of tsunami forecast or not,
(h) source of the heard warning, impression of warning upon hearing it,
(i) heard evacuation alert from LM or not,
(j) most beneficial source of information in the period from the earthquake to the day's sunset,

(k) intended doing evacuation before arrival of tsunami or not,
(l) evacuation place after the earthquake until sunset of the day, type of evacuation place,
(m) movement method, movement purpose,
(n) trigger for decision to start evacuation,
(o) reason for using a car for evacuation,
(p) problem of road for evacuation,
(q) problem of first evacuation place,

(r) watched hazard map beforehand or not,
(s) saw a board or a sign or a marking that indicated the direction and the place of evacuation, or not,
(t) made preparations for evacuation, such as securing furniture, preparing emergency carry-out bag, talking with family about tsunami emergency, pre-confirming evacuation place and road, checking tsunami hazard map, participating in community evacuation drill and so on, or not,
(u) knew the location of the designated place or building for evacuation near the place that was at the time of the earthquake or not,
(v) was able to go there or not,

(w) sex, age, job of interviewee, number of families living in the same house,
(x) saw the tsunami after the earthquake or not,
(y) damage to interviewee’s house by tsunami or by shaking of the earthquake,
(z) injury to interviewee and his/her family by tsunami or by shaking of the earthquake.

APPENDIX 2 Major trials for selecting explanatory variables (tried about 5 times of this table)

	R Correlation coefficient	R2 Adjusted determination coefficient	Significance level p of explanatory variable (Multiple regression analysis was applied to the set of variables of which cells are filled by p values).																			
			Start time of evacuation	Tsunami arrival time:*t*	Evacuation distance (here after, ED)	Walking ED	Car-using ED	Equivalent ED:*ed* 4)	Allowance period: *t/ed* 5)	Rate of car-using-evacuees: *cr*	Car velocity: *v*	Road serviceability: *v* × *cr* 6)	Rate of over 70 age persons	Tsunami warning height : *h* 7)	Rate of anticipating tsunami: *R*	*h* × *R*	Cognition rate of warning: *cr*	Warning effect *h* × *cr* 8)	Checked the hazard map or not	Participated evacuation drill or not	Preparedness 9)	Rate of returning home + detouring
1) Analyzed using the data of 19 local municipalities (here after LMs) of Iwate and Miyagi. Regression equation of polynomial terms was applied.	0.681	0.357	0.057											0.007					0.056			
	0.684	0.315	0.168										0.758	0.013					0.063			
	0.681	0.311	0.069								0.984			0.013					0.066			
	0.704	0.351	0.338		0.365									0.006					0.047			
	0.716	0.325	0.327		0.397					0.507				0.019					0.047			
	0.684	0.316	0.142											0.011	0.754				0.079			
	0.681	0.311	0.084											0.022					0.111	0.940		
	0.704	0.351	0.338			0.365								0.006					0.047			
	0.694	0.334				0.051					0.449			0.008					0.026			
	0.719	0.369				0.036						0.213		0.014					0.069			
	0.666	0.332				0.029								0.010					0.043			
	0.717	0.375				0.027					0.176			0.038					0.029			
	0.760	0.456						0.026			0.025			0.036							0.016	
	0.766	0.470				0.040						0.055		0.086							0.009	
	0.782	0.501					0.025					0.008		0.063							0.005	
	0.660	0.323										0.109		0.371							0.007	
	0.814	0.565						0.008				0.004		0.036							0.004	
	0.730	0.440						0.060				0.006									0.009	
	0.819	0.544						0.076				0.004	0.575	0.037							0.004	
	0.825	0.557						0.099				0.004		0.032	0.408						0.007	
	10) 0.814	0.532		0.971				0.044				0.007		0.059							0.005	
	0.817	0.539						0.013			0.665	0.083		0.051							0.005	
	0.760	0.456						0.026			0.025			0.036							0.016	
	0.797	0.530						0.015				0.004						0.068			0.013	
	0.644	0.248						0.415			0.049				0.872						0.052	
	0.730	0.440						0.060				0.006									0.009	
	0.744	0.426						0.003				0.007		0.176							0.061	
	0.650	0.306						0.009				0.025		0.122								
	0.793	0.523						0.003				0.003						0.127			0.031	
	0.748	0.471						0.008				0.003									0.012	
	0.744	0.425							0.012			0.038		0.544							0.041	
	0.804	0.545						0.002				0.003		0.087							0.009	
	0.748	0.433							0.023			0.015				0.461					0.041	
The following 2) and 3), same type of polynomial terms was applied.	0.842	0.596							0.040			0.011	0.021	0.126							0.003	
	0.842	0.596							0.139			0.002	0.023			0.125					0.003	
	0.810	0.523							0.108			0.005	0.062		0.605						0.007	
	0.805	0.513							0.077			0.005	0.080								0.010	0.950
	0.748	0.433							0.011			0.051		0.534							0.048	
	0.739	0.456							0.011			0.021									0.037	
	0.799	0.534							0.042			0.008	0.082								0.009	
	0.831	0.571							0.036			0.023	0.036	0.162							0.006	
2) Analyzed using the data of 17 LMs of Iwate and Miyagi.	0.803	0.568							0.013			0.020									0.006	
	0.822	0.577							0.010			0.065		0.280							0.006	
	0.838	0.603							0.078			0.010	0.509								0.004	
	0.831	0.619							0.031			0.008									0.003	
	0.850	0.631							0.024			0.027		0.258							0.003	
	0.861	0.656							0.057			0.004				0.148					0.009	
	0.836	0.599							0.009			0.009									0.004	0.558
3) 20 LMs of Iwate, Miyagi and Fukushima.	0.851	0.673							0.010			0.007									0.000	
	0.867	0.686							0.006			0.012		0.218							0.000	
	0.869	0.691							0.006			0.010						0.189			0.000	
	0.860	0.669							0.010			0.009					0.380				0.001	
Analyzed using the data of 20 LMs of Iwate, Miyagi and Fukushima. Monomial of exponential terms was applied.	0.823	0.617							0.001			0.002									11) 0.001	
	0.837	0.621							0.001		0.015						0.064				0.050	
	0.895	0.747							0.000		0.003			0.002							0.001	
	0.897	0.753							0.000			0.001		0.008							0.000	
	0.874	0.700							0.000			0.002					0.034				0.018	
	0.797	0.566							0.002									0.006			0.131	
	0.904	0.768							0.000			0.002						0.004			0.003	
	0.904	0.752							0.000			0.004	0.886					0.006			0.004	
	0.904	0.752							0.000			0.003			0.843			0.009			0.027	
	12) 0.827	0.599							0.001			12) 0.151						0.017			0.044	

1) The 19 LMs are Miyako, Yamada, Otsuchi, Kamaishi, Ofunato, Rikuzentakata, Kesennuma, Minamisanriku, Onagawa, Ishinomaki (rias coast area), Ishinomaki (flatland area), Higashimatsushima, Shichigahama, Miyagino-ku of Sendai, Wakabayashi-ku of Sendai, Natori, Iwanuma, Watari and Yamamoto.

2) Miyagino-ku and Iwanuma were omitted because of their large deviation of age distribution from Census. (Data of under 49 age of Miyagino-ku was 19%, whereas Census in the area was 57%. Data of over 70 age of Iwanuma was 5%, whereas Census was 18%.)

3) Shinchi, Minamisoma and Iwaki of Fukushima were added, whereas Soma of Fukushima was excluded because its preparation rate of emergency carry-out-bag was unnaturally low. (Preparation rate of emergency carry-out bag of Soma was 5.1%, while the rate of 20 LMs was 37.2% in average.)

4) Equivalent ED = Average walking ED × Rate of walking evacuee+Average car ED × Rate of car-using evacuee×0.244. (Here, ED is the distance of road from evacuee's home to the boundary of inundation area.)

5) Allowance period = Tsunami arrival time/Equivalent ED

6) Road serviceability = Rate of car-using evacuees × Car velocity (The better the road, the higher the car velocity and the higher the rate of car users.)

7) Tsunami forecast height broadcast by the first tsunami warning (Iwate 3m, Miyagi 6m, Fukushima 3m)

8) Warning effect = Tsunami forecast height × Cognition rate of tsunami warning

9) Preparedness= Rate of persons who prepared emergency take-out bag beforehand

10) Hereafter, HVI was modified by applying the rate of persons in their houses, as denoted by Eq. (2).

11) Hereafter, when calculating the preparedness rate, the denominator was replaced by the number of persons whose houses were completely collapsed to the number of persons who were at home or returned home before the tsunami's arrival and concurrently whose houses were completely collapsed.

12) Road serviceability was replaced by a dummy variable that defined the rias coast region (north of Oshika peninsula) as 0 and the flat coast region (southwest of Oshika peninsula) as 1.

REFERENCES

1) Imamura, A.: Disaster Prevention Committee Report 77., Disaster Prevention Committee, 1913. (in Japanese)
2) Coburn, A. and Spence, R.: *Earthquake Protection, 2nd Edition*, Chichester, John Wiley & Sons Ltd., 2002.
3) Ohta, Y., Goto, N. and Ohashi, H.: An Empirical Construction of Equations for Estimating Number of Victims at an Earthquake, *Zisin*, Vol. 36, No.6, pp. 463-466, 1983. (in Japanese)
4) Miyano, M. and Lu, H. G.: The Effect of Fault Distance to the Relationship between Human Casualties and Housing Damage due to Earthquakes, *Journal of Japan Society for Natural Disaster Science,* No.13-3, pp. 287-296, 1995. (in Japanese)
5) Murakami, H.: Chances of Occupant Survival and SAR Operation in the Buildings Collapsed by the 1995 Great Hanshin Earthquake, Japan, *Proceedings of 11th World Conference on Earthquake Engineering*, Paper No. 852, 1996.
6) Matsuda, I.: Changes of Human Losses and Deaths due to the 1995 Hyogo-ken Nanbu Earthquake, *Comprehensive Urban Studies,* Vol. 61 pp.155-166, 1996. (in Japanese)
7) Moroi, T. and Takemura, M.: Comparison of Fatality Risk between the 1891 Nobi Earthquake, the 1948 Fukui Earthquake and the 1995 Hyogoken-Nanbu Earthquake, *Zisin*, Vol. 52, pp.189-197, 1999. (in Japanese)
8) Takemura, M.: Earthquake and Disaster Prevention ~from vibration to aseismic design~, *Chukoshinsyo,* Chuokoron-Shinsha Inc., 2008. (in Japanese)
9) Kuwasawa, N., Kanai, M., Hosoi, K. and Katada, T.: Study on the effect of disaster education considering decision making evacuation from a tsunami, *Infrastructure Planning Review*, JSCE, Vol.23, pp.345-354, 2006. (in Japanese)
10) Suzuki, S. and Hayashi, H.: A Preliminary Critical Causal Analysis on the Mortality Caused by the 2011 East Japan Earthquake Tsunami, *Journal of Social Safety Science*, No.15, pp.179-188, 2011. (in Japanese)
11) Koyama, M., Ishii, N., Furukawa, A., Kiyono, J. and Yoshimura, A.: Municipality's Mortality Rate according to Inundation Level and Age Classes on the 2011 Great East Japan Earthquake, *Journal of JSCE A1 (Structure and Earthquake Engineering)*, Vol.69, No.4, I_161-I_170, 2013. (in Japanese)
12) Tanishita, M. and Asada, T.: Factors Associated with District Victim Rate by Tohoku Earthquake Tsunami in Minamisanriku Town, *Journal of JSCE A1 (Structure and Earthquake engineering).* Vol.70, No.4, I_66-I_70, 2014. (in Japanese)
13) Goto, Y.: Index to Evaluate Tsunami Evacuation Potential and its Validation at Yamada, Iwate Prefecture, *Journal of Disaster Research*, Vol.9, No.7, pp.730-742, 2014.
14) Nakasu, T., Ono Y. and Pothisiri W.: Forensic Investigation of the 2011 Great East Japan Earthquake and Tsunami disaster: a case study of Rikuzentakata, *Disaster Prevention and Management: An International Journal*, Vol. 26, Issue 3, DOI: 10.1108/DPM-10-2016-0213, 2017.
15) Nakasu, T., Ono Y. and Pothisiri W.: Why did Rikuzentakata have a high death toll in the 2011 Great East Japan Earthquake and Tsunami disaster?: Finding the devastating disaster's root causes, *International Journal of Disaster Risk Reduction*, Vol.27, pp.21-36, 2017.
16) Ministry of Land, Infrastructure, Transport and Tourism: Summary of Tsunami Affected Urban Area Reconstruction Method Investigation Survey from the Great East Japan Earthquake http://www.mlit.go.jp/toshi/toshi-hukkou-arkaibu.html. (in Japanese, last accessed in February, 2015)
17) City Bureau of Ministry of Land, Infrastructure, Transport and Tourism: About the Arrangement of Escape Routes, Evacuation facilities and Evacuation Guidance assuming Tsunami Evacuation (Third Edition) 2013, http://www.mlit.go.jp/common/000233464.pdf. (in Japanese, last accessed in June, 2016)
18) Goto, Y. and Nakasu, T.: Human Vulnerability Index for Evaluating Tsunami Evacuation Capability of Communities, *Proceedings of 16th World Conference on Earthquake Engineering*, Paper No.1669, 2017.
19) Iwate Prefecture: Iwate Disaster Management Portal - Human Damage / Building Damage Status due to Tohoku Region Pacific Offshore Earthquake Summary as of July 31, http://www2.pref.iwate.jp/~bousai/index.html. (in Japanese, last accessed in September, 2015)
20) Miyagi Prefecture: Damage caused by the Great East Japan Earthquake as of July 31, 2015 http://www.pref.miyagi.jp/uploaded/attachment/321498.pdf. (in Japanese, last accessed in September, 2015)
21) Fukushima Headquarter for Disaster Countermeasures: Damage Status due to the 2011 Tohoku Region Pacific Offshore Earthquake (1666th Report) as of October 31, https://www.pref.fukushima.lg.jp/uploaded/life/241248_559150_misc.pdf. (in Japanese, last accessed in October, 2016)
22) City Bureau of Ministry of Land, Infrastructure, Transport and Tourism of Japan and the Center for Spatial Information Science of the University of Tokyo: Fukkou Shien Chosa archive. http://fukkou.csis.u-tokyo.ac.jp/default/about. (in Japanese, last accessed in March, 2014)

23) Goto, Y., Ikeda, H., Ichiko, T., Ogawa, Y., Kitaura, M., Sato, S., Suzuki, H., Tanaka, T., Nakamura, M., Mikami, T., Murakami, H., Yanagihara, S. and Yamamoto, K.: The Joint Survey Group about the Tsunami Evacuation of the Great East Japan Earthquake and Field Study by the Sub-group charged with Yamada and Ishinomaki – Analysis on Data Characteristics – , *Journal of JAEE*, Vol. 15 (2015) No. 5, p. 5_118-5_143, 2015. (in Japanese)

24) Statistics Bureau of Ministry of Internal Affairs and Communications: Statistics Viewed on the Map (Statistical GIS)
http://e-stat.go.jp/SG2/eStatGIS/page/download.html. (in Japanese, last accessed in October, 2016)

25) Goto, Y.: Impacts of Returning Home or Dropping In Just Before or During Evacuation from the 2011 East Japan Great Earthquake Tsunami, *Journal of JAEE*, Vol. 16 (2016) No.10, p.10_86-10_104, 2016. (in Japanese)

26) Mikami, T.: The Survey Analysis about the Victims by the Tsunami under the Great East Japan Earthquake – Yamada-cho and Ishinomaki City –, *Journal of JSCE A1(Structure and Earthquake engineering)*, Vol.70, No.4, I_908-I_915, 2014. (in Japanese)

27) Tanaka, Y. and Tarumi T.: Statistical Analysis Handbook - Multivariate analysis, ISBN 4-320-01492-8, Kyoritsu Shuppan Co., Ltd., pp.48-49, 1995. (in Japanese)

28) Tanishita, M.: Factors associated with district tsunami victim rate by age class in Ishinomaki city 2011 Great East Japan earthquake and tsunami, *ResearchGate*, DOI: 10.13140/RG.2.1.1493.8002/1, 2016.
https://www.researchgate.net/publication/299752194. (in Japanese, last accessed in November, 2016)

29) Koshimura, S. and Gokon, H.: Structural Vulnerability and Tsunami Fragility Curves from the 2011 Tohoku Earthquake Tsunami Disaster, *Journal of JSCE B2 (Coastal engineering)*, Vol.68, No.2, I_336-I_340, 2012. (in Japanese)

30) Japan Meteorological Agency: On the Direction of Improvement of the Tsunami Warning based on the Damage of the Tsunami caused by the Tohoku Region Pacific Offshore Earthquake, September 12, 2011, http://www.jma.go.jp/jma/press/1109/12a/torimatome.pdf. (in Japanese, last accessed in October, 2016)

31) Sekimoto, Y., Nishizawa, A., Yamada, H., Shibasaki, R., Kumagai, J., Kashiyama, T., Sagara, T., Kayama, Y. and Ootomo, S.: Data Mobilization by Digital Archiving of the Great East Japan Earthquake Survey, *Theory and Applications of GIS*, GIS Association of Japan, Vol.21, No.2, pp.1-9, 2013.

(Original Japanese Paper Published: May, 2017)
(English Version Submitted: June, 4, 2018)
(English Version Accepted: August 22, 2018)