International Institute for Applied Systems Analysis
IIASA www.iiasa.ac.at

Address: IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria

Email: permissions@iiasa.ac.at

Young Scientists Summer Program

Measuring global social vulnerability to natural hazards at the subnational level

Xueting Li, lixt021@gmail.com
Raya Muttarak, muttarak@iiasa.ac.at
Roman Hoffmann, hoffmannr@iiasa.ac.at

Approved by:

Mentor(s): Raya Muttarak, Roman Hoffmann
Program: POPJUS
Date: 30.09.2022

This report represents the work completed by the author during the IIASA Young Scientists Summer Program (YSSP) with approval from the YSSP mentor.

It was finished by _Xueting Li, Raya Muttarak and Roman Hoffmann_ and has not been altered or revised since.

Mentor signature:

Table of contents

ZVR 524808900

Disclaimer, funding acknowledgment, and copyright information:
IIASA Reports report on research carried out at IIASA and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the institute, its National Member Organizations, or other organizations supporting the work.

UPDATE OR DELETE
For IIASA Working Papers/Reports/Policy Briefs funded by the IIASA core budget we ask the authors to include the following:
The authors gratefully acknowledge funding from IIASA and the National Member Organizations that support the institute (The Austrian Academy of Sciences; The Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES); The National Natural Science Foundation of China (NSFC); The Academy of Scientific Research and Technology (ASRT), Egypt; The Finnish Committee for IIASA; The Association for the Advancement of IIASA, Germany; The Technology Information, Forecasting and Assessment Council (TIFAC), India; The Indonesian National Committee for IIASA; The Iran National Science Foundation (INSF); The Israel Committee for IIASA; The Japan Committee for IIASA; The National Research Foundation of Korea (NRF); The Mexican National Committee for IIASA; The Research Council of Norway (RCN); The Russian Academy of Sciences (RAS); Ministry of Education, Science, Research and Sport, Slovakia; The National Research Foundation (NRF), South Africa; The Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS); The Ukrainian Academy of Sciences; The Research Councils of the UK; The National Academy of Sciences (NAS), USA; The Vietnam Academy of Science and Technology (VAST).

For research funded by an external third party we ask the authors to include the full name of the funder, the title of the project and the grant number, e.g.
The authors gratefully acknowledge funding from the European Research Council for the research project 'Forecasting Societies Adaptive Capacities to Climate Change' (FUTURESOC, FP7 230195).

Abstract

Identifying and addressing social vulnerability is an integral element of disaster risk reduction efforts. A range of quantitative approaches have been proposed to measure vulnerability in different contexts, guiding the scale of research and the selection of indicators. To fully understand the interplay between hazard, exposure, and vulnerability, the approach of social vulnerability measurement must be validated against the actual impacts of natural hazards in various locales. However, studies that are highly localized have difficulty capturing the common characteristics of societies and comparable factors that shape social vulnerability at a larger scale, making global comparisons impossible. This study will investigate global social vulnerability to natural hazards using subnational data for a large number of countries. Various sources of demographic and socio-economic data were employed to construct a social vulnerability index (SVI). The index was validated with actual impact of natural disasters using a second dataset. Results show that socio-economic development and population structure mainly summarize social vulnerability characteristics, while education has the strongest negative effect on total death. Further investigations in identifying subnational subgroups and construct a second set of SVI incorporating more data is required.

About the author

Xueting Li is now Ph.D. student in Asian Demographic Research Institute, Shanghai University. (Contact: lixt021@gmail.com)
Raya Muttarak is the former IIASA Population and Just Societies (POPJUS) Program director and Principal Research Scholar in the Migration and Sustainable Development (MIG) Research Group. She is currently professor of Demography and the Department of Statistical Sciences at the University of Bologna. (Contact: muttarak@iiasa.ac.at)
Roman Hoffmann is the research group leader of the Social Cohesion, Health, and Wellbeing (SHAW) Research Group in the Population and Just Societies Program in IIASA. (Contact: hoffmannr@iiasa.ac.at)

Acknowledgments

This research is funded by National Natural Science Foundation of China. Xueting Li gratefully acknowledges support from IIASA and ADRI. In addition to the help and insightful feedback provided by Roman, Raya and Prof. Leiwen Jiang, I would like to extend the appreciation for the support and encouragement from POPJUS.

1. Introduction

From 2000 to 2019, the number of reported natural disasters increased more than 1.5 times by comparison with 1990 to 2000. More than 90% of these reported disasters were climate-related, accounting for approximately 0.5 million deaths and 3.9 billion people affected. In the same period, the economic losses associated with disasters were growing worldwide(CRED & UNDRR, 2020). The increasing frequency of climate change-induced extreme weather events has generated serious effects on the well-beings of human and requires greater strengthening of climate-related risk reduction efforts. Accordingly, climate action is one of the priorities of the UN Sustainable Development Goals (SDGs) which explicitly promote adaptive responses for reducing adverse climatic impact (United Nations, 2015).

To this end, the concept of risk is fundamental to understand the impact of climate change. Risk is commonly regarded as a function of hazards, exposure and vulnerability, each may be affected by the spatial and temporal change of socio-economic context (Reisinger et al., 2020). According to the Intergovernmental Panel on Climate Change (IPCC) in its Fifth Assessment Report, vulnerability is defined as "the likelihood to be affected, sensitivity to be harmed and lack of coping ability" (IPCC, 2012). Given that exposure and the impact of natural hazards is not distributed evenly across population subgroups and locations, the concept of social vulnerability has been proposed in recognizing that vulnerability goes beyond physical characteristics (Adger, 1999; Cutter et al., 2003). In general, social vulnerability refers to the characteristics of a person or a group in terms of their capacity to anticipate, cope with, resist and recover from the impact of a natural hazard (Wisner et al., 2003). It is a multi-dimensional concept which incorporates the demographic, societal, economic, political aspects of the society.

Due to different research framework and pre-analytic condition, there has not been a consensus on the major factors determining social vulnerability. One of the most referred models, the hazards-of-place model, was developed to measure the regional social vulnerability to natural hazards in the U.S., by constructing a social vulnerability index(SoVI) using a set of socioeconomic and demographic variables (Cutter et al., 2003). However, the measurement of social vulnerability is available only for one specific or a few countries (e.g. the US, Germany, India…) (Alexander Fekete; Flanagan, Hallisey, et al., 2018; Maiti et al., 2015). Existing vulnerability indices that cover several countries are available at a country level(Katharine, 2004) . Presenting an index at a large geographical scale is not particularly meaningful for the identification of the most vulnerable subgroups and their location for interventions and allocation of resources to mitigate climatic risk (Wood et al., 2021)..

Furthermore, often, the indicators adopted to measure social vulnerability are different, creating difficulties in capturing the common characteristics of societies and comparable factors shaping social vulnerability at a larger scale, challenging comparisons at the global level(Tate, 2012). Likewise, existing social vulnerability indices are rarely validated and hence it is not clear whether they actually capture the vulnerability underlying climatic risks. Typically, the lack of empirical data constrains the validation of the constructed social vulnerability indices (e.g. with the impact of actual hazards) (A. Fekete, 2009; Rufat, Tate, Burton, & Maroof, 2015).

This research aims to close the research gaps by: 1) measuring global social vulnerability to natural hazards using subnational data for a large number of countries; and 2) validating the constructed social vulnerability index against the actual economic losses and fatalities from natural hazards. Various sources of demographic and socio-economic data were employed to construct a social vulnerability index (SVI). The index was validated with actual impact of natural disasters using a second dataset, the Emergency Events Database (EM-DAT). Principle Component Analysis and Poisson regression model were adopted to develop the SVI and to determine the contributions of different socio-economic factors to social vulnerability.

Results show that socio-economic development and demographic structure mainly explain the underlying social vulnerability. Demographic, human capital and economic development have a significant effect on disaster fatality, while education has the strongest negative effect on total disaster death. Model results show that higher proportion of women or/and older populations do not necessarily lead to higher social vulnerability in a particular region, while life expectancy significantly has a negative effect. Subnational regions in middle and southern Africa and southern South America were more socially vulnerable to natural hazards during the studied periods (add years of study) while, eastern China became more vulnerable over time.

The remainder of this report is structured as follows. The next section will focus on the theoretical background. Section 3 will go through data sources and methodology. Section 4 will present the model result as well as explore the spatial patterns of social vulnerability, followed by a discussion and conclusion in the last section.

2. Theoretical framework and previous literature

2.1 Social vulnerability frameworks

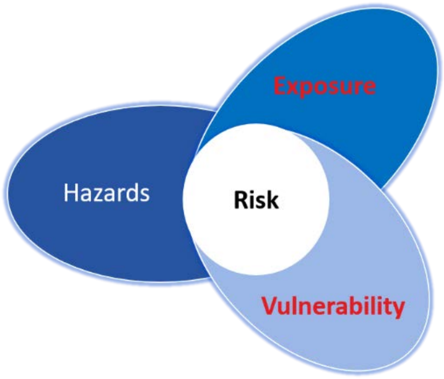

Figure 1. the conceptual framework of risk (IPCC, 2014)

Note: Risk of climate-related impacts results from the interaction of climate-related hazards (including hazardous events and trends) with the vulnerability and exposure of human and natural systems.

In the conceptual framework of risk (Fig.1), vulnerability is generally defined as the potential to suffer harm or loss following the IPCC's definition. Social vulnerability accounts for the source of vulnerability to disasters from a social perspective (Burton et al., 2018). In general, social vulnerability refers to the social and demographic characteristics of human population that determines the (in)ability of individuals or groups to withstand adverse impacts from or cope with external hazards to which they are exposed (Cutter et al., 2003; Cutter & Finch, 2008).

There are three classic conceptual framing of social vulnerability including risk-hazard, political ecology, and social-ecological system frameworks (Eakin & Luers, 2006). The risk-hazard framework is represented by the Human Adjustment to Natural Hazards model which regards a hazard as the result of the interaction between a natural event and human activity. Overall, critics argued that this approach may poorly describe the contribution of human society to hazards (Turner et al., 2003). While the risk-hazard perspective views vulnerability mostly as function of natural events and managerial decisions, the disaster Pressure and Release Model (PAR) from the political ecology perspective considers social vulnerability as a sequence of root causes, dynamic pressures and unsafe conditions. These frameworks face the difficulty of lacking consideration on natural-human systems interactions (Burton et al., 2018).

Scholars who measure the risk of disasters through a social vulnerability lens consider vulnerability to be an essential element of disaster risk. People face different degrees of disaster risks because of their various sociodemographic characteristics. The Hazards of Place Model addresses social vulnerability from both biophysical and social systems to a specific place and time(Cutter, 1996). This approach provides a useful framework for identifying socially vulnerable population groups, projecting regional disaster risks, and preparing for the occurrence of events (Rufat et al., 2019). Moreover, by using census data, this approach allows a consistent set of variables that can be monitored over time and space. There have been many studies carried out to investigate social vulnerability to natural disasters in both developed and developing countries using this approach (Frigerio & Amicis, 2016; Siagian et al., 2014; Zhou et al., 2014).

2.2 Quantitative measurement of social vulnerability

Due to different research framework and pre-analytic conditions, there has not been a consensus on the major factors for determining social vulnerability. The initial usage of social vulnerability indicators is derived from social science research which adopted social indicators to measure quality-of-life or human development (Zhou et al., 2014) . The social vulnerability index (SVI) developed by Cutter et al. (2003) and others is usually composed of a set of socio-economic, demographic and other characteristics of population. Recently considering the continued cultivating effects of the cascading disasters in recent decades, the Geospatial Research, Analysis, and Services Program (GRASP) at Centers for Disease Control and Prevention (CDC) /Agency for Toxic Substances and Disease Registry (ATSDR) in the U.S. created a Social Vulnerability Index (SVI) database drawing on the hazard of place model (Flanagan, Hallisy, et al., 2018). This tool is designed to identify and map socially vulnerable populations at various community scales (Wolkin et al., 2015) . The approach would facilitate social vulnerability measurement at different spatial scales, thus making data compilation more efficient to craft targeted solutions and strategies to improving local adaptive capacity and resilience.

Commonly used variables to capture social vulnerability are gender, age, and income status (Cutter et al., 2003).,Women, children, the elderly, and population with low socio-economic status are recognized as being particularly affected when a disaster strikes and need more support (Rufat, Tate, Burton, & Maroof, 2015). In addition to the commonly used variables mentioned above, there are other socio-economic factors that are important for assessing social vulnerability. For instance, education is one of the important socio-economic factors which shape people's response and adaptive capacity to disaster risks. Such factor can affect people's ability to access to information and resources for coping with disasters before, during, and after an event. Socio-economic factors can even lead to different environmental exposures by affecting people's choice of residence. Moreover, quality of human settlement and built environment including health condition are are important in risk identification process as well.

2.3 Global social vulnerability assessments

The selection of social vulnerability indicators mostly depends on data availability and the social context in which vulnerability occurs. Although it is possible that the same factor contributes differently to social vulnerability in different contexts, several factors are identified as having a similar significant effect on the overall social vulnerability. For example, a German case study found that age played a key role in people's ability to obtaining information and receiving help during flood. This research also shares the founding with an Italian study of the effect of occupation or socio-economic status on people's prior awareness before disaster (Kuhlicke et al., 2011).

Moreover, different studies focusing on social vulnerability to a certain type of natural hazards identified a few common characteristics of vulnerable population. For example, health is one of the internal factors of social vulnerability to both flood and drought. Health is related to income, education and other socio-economic factors whereby health inequality may aggravate the health condition of people in low socio-economic status (Otto et al., 2017). Also, studies focusing on social vulnerability in urban area or the overall vulnerability both take housing quality into account, as poorer housing condition may more likely to be adversely impacted by disasters (Davino et al., 2021; Mallick et al., 2011).

An overall social vulnerability is determined by several factors. Inductive approaches used to quantitatively measure social vulnerability not only identify the confounding factors but also consider their changing trends. The purpose of aggregating the factors is to compare the overall differences of social vulnerability (Rufat, Tate, Burton, & Maroof, 2015). Most social vulnerability studies adopting the inductive approach were also carried out for one specific location, such as a city, a country or a pre-defined region. There is a limited number of studies assessing regional social vulnerability from a global perspective. Fast changing social fabric led by urbanization and industrialization worldwide, and its global consequences are not yet well understood in the context of climate change (UNU-EHS, 2008).

3. Methods and data

3.1 Data sources

Global Data Lab: Subnational demographic and socio-economic data are obtained from the Area Database of the Global Data Lab (GDL). This Area Database provides publicly available subnational development indicators for low- and middle-income countries (LMICs). The indicators are constructed by aggregation from representative household-level datasets which were harmonized and combined into one database(Smits & Permanyer, 2019) . This aggregated data allows to study the subnational demographic and socio-economic disparities within country and enable comparison among different countries. The major data source of this database are census and surveys which have large samples, such as Demographic and Health Surveys, UNICEF Multiple Indicator Cluster Surveys and IPUMS International. Other country specific surveys are also included for the country-year where these sources are not available. The database currently contains over 100 subnational indicators for 1400+ regions across LMICs countries between 1990 and 2020 (Global Data Lab).

EM-DAT data: The impact data of natural hazards are obtained from the Emergency Events Database (EM-DAT), which is managed by the Centre for Research on the Epidemiology of Disasters (CRED). The EM-DAT database contains the occurrence and effects of natural, technological and complex disasters from 1900 to the present day. The disaster classification used in EM-DAT is based on and adapted from the he IRDR Peril Classification and Hazard Glossary. It contains information on geophysical, meteorological, hydrological, climatological, biological and extra-terrestrial disasters. A disaster event is included in the database if it fulfills at least one of the following criteria: (1) Ten or more reported deaths. (2) One hundred or more reported affected population. (3) Declaration of a state of emergency or appeal for international assistance (EM-DAT). The database contains human impact, economic impact, sectorial impact and infrastructural impact information of a disaster event.

GDIS disaster data: The geocoded information of the natural disaster events is obtained from the Geocoded Disasters (GDIS) database, which is the extension of the EM-DAT database. It provides subnational locations for natural disasters recorded in EM-DAT between 1960 and 2018. The highest spatial resolution of the locations in this database corresponds to administrative level 3 (usually district/commune/village) in the Global Administrative Areas database (Rosvold & Buhaug, 2021).

3.2 Measuring disaster impacts

Considering the data availability, we use information on deaths and missing persons as variables to measure the impact of a disaster. Data on disaster events was retrieved from the EM-DAT database. The variable deaths refer to the number of people who lost their life in a disaster event. Missing persons refers to the number of people whose whereabouts since the disaster are unknown. As these are often presumed dead by the authorities, the missing persons are counted as part of the total deaths of a disaster, the main impact variable we are considering in our analysis.

In order to account for the overall exposure of the population to a hazard and to isolate the role of vulnerability in shaping disaster outcomes, we control for the total population affected by a disaster in our models. This measure is distinct from the impact measure described above. Affected refers to people requiring immediate assistance during an emergency situation. There are different definitions and criteria to estimate the count of the people being affected, which is broadly defined as the sum of all injured, homeless, and otherwise affected persons.

The analysis was carried out at the subnational regional level. However, disasters often affect several regions and the EM-DAT data provides information only in an aggregated form at the country level. If a disaster had affected several regions, we distributed both the total deaths and the number affected for each region according to their relative population size.

While this entails a set of assumptions, it was a necessary step to enable us to perform the subsequent subnational analyses at the global level.

3.3 Measuring socioeconomic and demographic vulnerability

3.3.1 Selection of indicators

Previous studies consistently found that lacking access to resources and services, limited social capital, and inadequate building environment are among the main factors influencing social vulnerability (Adger & Kelly, 1999; Sapam et al., 2014). However, there has always been a disagreement concerning the selection of variables that represent and measure these broad concepts. Based on social vulnerability indicators proposed by the previous literature and considering data availability this study examines the global social vulnerability at the subnational level adopting nine indicators to construct the SVI. These indicators cover not only the generally accepted contributors to vulnerability such age, gender, and socioeconomic status but also take health into account.

We used interpolation and imputation methods (bootstrapping and the EM algorithm) to impute missing values for the selected indicators. The algorithm uses the familiar EM (expectation-maximization) algorithm on multiple bootstrapped samples of the original incomplete data to draw values of the complete-data parameters. The algorithm then draws imputed values from each set of bootstrapped parameters, replacing the missing values with these draws (Honaker et al., 2011). Following the imputation, we aggregated the data to five year's average values (last one is three years due to data availability) for each sub-national region between 2000 and 2018. The nine indicators are divided into four main aspects (demographic; human capital; housing condition; economic development) that determine social vulnerability of a region. The indicator list is shown in Table 1 below.

Table 1. The indicator list of the SVI

No		Variable	Description
1	Demographic	young	Youth dependency ratio: the young (<15) compared to the working age population (15-64)
2		old	Old age dependency ratio: the old (>64) compared to the working age populations (15-64)
3		women	% of women aged 50+
4		hsize	Average household size
5	Human capital and health	edu	Mean years of education of adults aged 20+
6		lifexp	Life expectancy at Birth
7	Housing condition	elec	% households with electricity
8	Economic development	wealth	Mean International Wealth Index (IWI) score
9		urban	% of population living in urban areas

3.3.2 Rationale for choosing social vulnerability indicators

Demographic characteristics: Children and the elderly are among the demographic groups most affected by natural hazards (Wisner et al., 2003). Children are vulnerable during a hazard due to lack of knowledge, life experiences and knowledge to protect themselves. Older people are also vulnerable due to limited access to resource. underlying health conditions, and lack of mobility. High proportion of children and older people may increase the burden of care and decrease the resilience of a household. From the literature, we can assume that the greater proportion of children and older people make the region more vulnerable to cope with and recover from a disaster.

There are studies arguing that larger households are in a stronger position than smaller ones in absorbing shocks and diversifying environmental risks. However, there is also evidence showing that despite of the shrinking average household sizes, family networks still provides basic and fundamental support for each individual in the time of socioeconomic and environmental shocks. Compared to men, women are considered as more socially vulnerable to natural hazards because of their often disadvantaged role in society (Fothergill, 1996). In labor markets, for instance, women may have less employment opportunities (Flatø et al., 2017; Ma, 2016) and when experiencing shocks having less financial and social resources to cope with the shocks. The situation might be more severe when a disaster strikes. Therefore, youth dependency ratio, old dependency ratio, household size and the proportion of women are adopted to capture the demographic characteristics of a region.

Human capital and health: Human capital is another important factor to be taken into consideration in analyzing social vulnerability. Educational attainment not only implies employment opportunities, but also elevated quality of life(Lutz et al., 2014). Higher levels of education improve the ability to comprehend warning information, prepare for onset of environmental changes, and have access to resources for coping with the situation and recovering from the damage (Hoffmann & Muttarak, 2017; O'Neill et al., 2020). Also, evidence indicates people with poor health condition are more vulnerable at all stages of a disaster event (Flanagan, Hallisy, et al., 2018). Thus, this study uses mean educated years, life expectancy as proxy of health status, to measure the potential effects of education and health on social vulnerability.

Economic development: Socioeconomic development enables urban regions to quickly absorb and recover from disasters, but it also means that a larger population and economic assets are under potential risk(Garschagen & Romero-Lankao, 2015). Likewise, generally less materials and financial resources contribute to higher social vulnerability to hazards(Rufat, Tate, Burton, & Maroofb, 2015). The international wealth index is a broadly used indicator constructed by the Global Data Lab, which measures the standard of living of households across the developing world on the basis of ownership of consumer durables, housing characteristics and access to basic services (Global Data Lab). Both the proportion of urban population in a region and the international wealth index are used to measure the status of economic development of a subnational region.

Housing conditions: Housing conditions are also a significant aspect of social vulnerability(Ignacio et al., 2016). Environmental displacement of populations is more common in rural areas than urban ones largely because of the differences in housing and living conditions(Rodenbiker, 2020). Therefore, proportion of households with electricity is included to measure the potential effects of housing condition on social vulnerability.

3.4 Analytical framework and estimation

PCA is adopted for constructing of the social vulnerability index. This data-driven approach derives demographic and socio-economic characteristics that summarize social vulnerability profiles. The demographic and socio-economic characteristic of vulnerable population underlying regional capacity to cope and recover capacity from hazards. It is not enough to only delineate the pre-existed condition of society by constructing the SVI, explore how it related to actual consequences of natural hazards is necessary. To validate the constructed social vulnerability indicator, model-driven approach was also adopted by this study. Poisson regression model is used to explore the relationship between social vulnerability characteristics with actual impact from natural hazards. Therefore, the set of demographic and socio-economic variables is first analyzed by PCA and then validated by means of a multivariate Poisson regression model.

3.4.1 Principal component analysis

Following the approach pioneered by Cutter et al.(Cutter et al., 2003) , the first set of the SVI was constructed by PCA. This inductive approach decomposes an extensive set of variables into a limited number of uncorrelated principle components that explain most of the variance of the original dataset(Burton et al., 2018). The larger the variance a principle component explains, the more information this dimension contains for measuring social vulnerability. This method helps to identify a robust and consistent set of variables that make a region socially vulnerable to natural hazards over the years.

In total we utilized nine variables to construct the global SVI at the subnational level for four periods of 2000-2004, 2005-2009, 2010-2014 and 2015-2018 respectively.

The SVI construction using PCA involves three steps: 1) variable normalization; 2) principle component analysis and; 3 the composite social vulnerability index. Firstly, the input variables were normalized to z-scores, each with mean zero and standard deviation of one so as to ensure the variables are comparable to each other. Then, Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure were performed to examine the suitability of the data for PCA. Next, PCA was performed in R software using varimax rotation and Kaiser criterion (eigenvalues greater than 1) to extract a set of components representing social vulnerability. The factor score of each component is placed in an additive model to produce the composite social vulnerability index score (SVI) for each subnational region. The factor scores of each component implicitly assume weights of each variable. Because there is no theoretical basis for assuming one factor is more significant than another one in the construction of the index, following Cutter et al., this study uses sum of the factor scores to produce the composite SVI in each period. A positive sign is assigned to a factor score which increase the total vulnerability, and a negative score is assigned if it decreases the vulnerability. In this way, the higher SVI, the more vulnerable a region is in the referred period.

3.4.2 Model based approach: Poisson regression

We performed regression modeling of fatality from natural hazards while controlling for the affected population. Because the outcomes are non-negative count variables, the Poisson regression model was adopted to regress the actual impact of natural hazards on the different demographic and socio-economic characteristics of the regions.

$$\ln(y_{it}) = \beta_0 + X_{it}\beta + \varepsilon_{it}$$

In the Poisson model, each unit *i* corresponds to a subnational region in which y_{it} events are observed during a time window *t*. The dependent variable y_{it} is the number of deaths normalized by the population size of the region *i*. The estimates β show the estimated percent change in the expected fatality for a one unit change in the demographic and socio-economic characteristics.

4. Results

4.1 Descriptive statistics

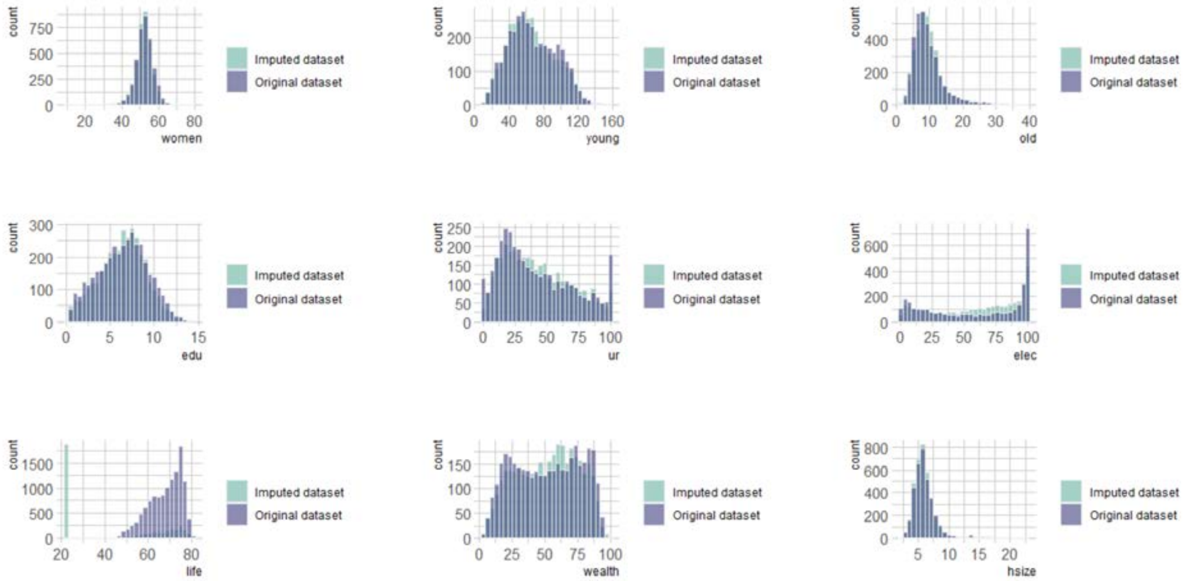

Figure 2. The distribution of each variable before and after imputation

Table 2. Variable description of the imputed dataset

No	Variable	Range (%)	Mean(sd)
1	women	17.88~69.47	51.96±4.43
2	edu	0.03~13.96	6.32±2.62
3	wealth	2.25~98.36	52.42±23.24
4	young	10.72~150.98	65.17±27.37
5	old	1.14~37.99	9.75±4.41
6	hsize	2.31~10.45	5.89±1.69
7	elec	0.21~100	62.45±32.82
8	life	23~79.83	42.06±22.79
9	ur	0~100	42.76±26.37

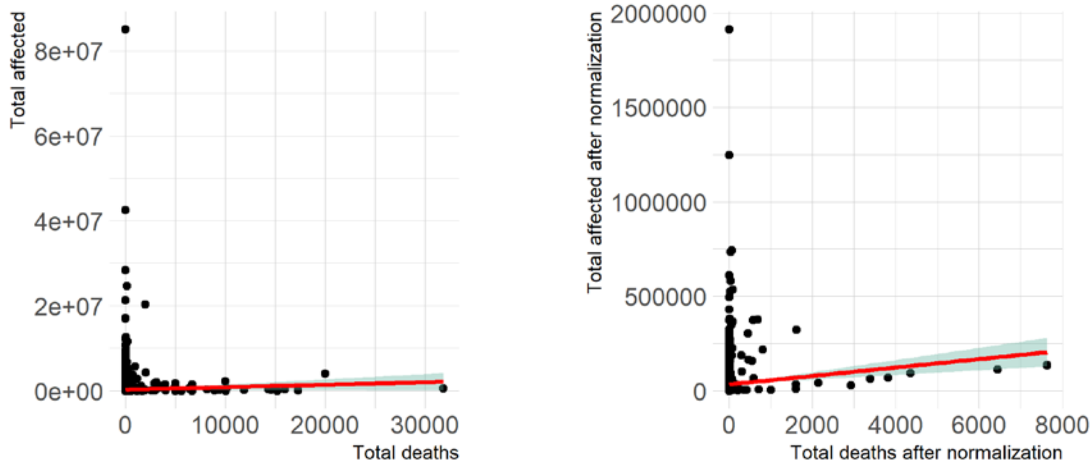

Figure 3. The distribution of total deaths and affected population before and after normalization

Fig.2 shows that the distribution of each predict variable in the imputed dataset is in accordance with the original dataset. Although the life expectancy variable in the imputed dataset clustered around low value, it may due to the amount of missing value during the first research period. Table 2 shows the mean and standard deviation of the imputed dataset. The proportion of women is slightly higher than half the population aged 50+, also is the international wealth index and the housing condition indicator. There are some observations with relatively high youth dependency ratio and old age dependency ratio. Likewise, large differences in the urbanization rate are observable across regions.

Fig 3 shows clearer relationship between total deaths and affected population after normalization. As expected, the figures are highly skewed and not normally distributed.

4.2 Principal component analysis

Table 3. Principle components of the SVI

No	Principle component	Driving indicators	Loadings
PC1	Socio-economic development Proportion variance: 0.46	elec	0.91
		wealth	0.9
		young	-0.79
		edu	0.78
		ur	0.74
PC2	Population structure Proportion variance: 0.20	old	0.76
		women	0.74

Several tests were carried out before performing the PCA for each period. The values of KMO of the matrixes are all greater than 0.7, and the results of Bartlett's test of sphericity are significant. It suggests that the datasets are appropriate for performing principle component analysis. Using the Kaiser criterion, two principle components explain about 66% of the variance of the total nine social vulnerability indicators in each period (Table.2). The underlying dimensions of social vulnerability remain unchanged in the period, of which the drivers are identified with absolute loadings greater or equal to 0.7.

The dominant drivers of social vulnerability of the first principle component are housing condition (proportion of households with electricity), demographic characteristics (youth dependency ratio,), human capital (average educated years), economic development (international wealth index, proportion of urban population), the total variation explained is about 46%. The second principle component is mostly driven by demographic characteristics (old age dependency ratio, % women aged 50+), the total variation explained is about 20%.

The first principle component is a major contributor to social vulnerability because of its highest explained variance in each decade. PC1 characterizes regions with more households with electricity, higher wealth, less youth dependency ratio, higher level of education and more urban population. PC2 characterizes regions with higher old age dependency ratio and more women aged 50+. The first principle component mainly captures regional characteristics of socio-economic development, and the second one delineates population structure. Thus, the first PC is assigned negative for indicating regions of lower social vulnerability, the second is assigned positive cardinality for indicating regions of higher social vulnerability. The SVI is calculated as follows:
*SVI = -PC1 + PC2

4.3 Modeling vulnerability drivers

Table 4. Poisson regression model result

	Poisson regression	Linear regression
	deaths/pop	deaths/pop (log)
affected pop/	0.00000***	0.048
affected pop(log)	(0.000)	(0.043)
	-0.040***	
young	(0.0002)	
	0.043***	
old	(0.001)	
	0.083***	
women	(0.001)	
	-0.012***	
hsize	(0.002)	
	-0.373***	
edu	(0.003)	
	-0.005***	
lifexp	(0.0004)	
	0.001	
elec	(0.0002)	
	-0.034***	
wealth	(0.0004)	
	0.002***	
urban	(0.0002)	
period	0.945***	0.192
"2005-2009"	(0.010)	(0.265)
period	1.475***	0.307
"2010-2014"	(0.018)	(0.265)

period "2015-2018"	-0.791*** (0.027)	-0.946*** (0.310)
PC1		-0.251** (0.101)
PC2		-0.352*** (0.098)
AIC	582152.4	
R2		0.016

Note: *p<0.1; **p<0.05; ***p<0.01, n = 2532

Observations in different period were pooled together, in total there are 2532 subnational locations with actual impact record included in the model. Controlled the affected population normalized by total population size, Poisson regression model results show that the set of demographic and socio-economic variables except the housing condition indicator (proportion of households with electricity) all have significant effects on disaster fatality. After exponentiation transformation, each one-unit increase in old age dependency ratio, expected fatality increases by 4.4%. But for life expectancy, disaster fatality is expected to decrease -0.5%. Each one-unit increase in proportion of women aged 50+ also leads to expected fatality increases 8.6%. Education level has the strongest effect on disaster fatality. One-unit increase in average educated years associates with 31.1% decrease in disaster fatality. Wealth status is negatively associated with disaster fatality. Each one-unit increase of it associated with expected fatality decrease 3.4%. Youth dependency ratio and household size are both negatively associated with disaster fatality, these two indicator are collinear with the wealth indicator. Fr one-unit increase of the urbanization level, disaster fatality is expected to have an increase of 0.2%. This may be related to high population density in urban area. In all, education has the strongest effect on disaster fatality, also population age and gender structure play important role in determine impact of disasters.

4.4 Mapping vulnerability globally

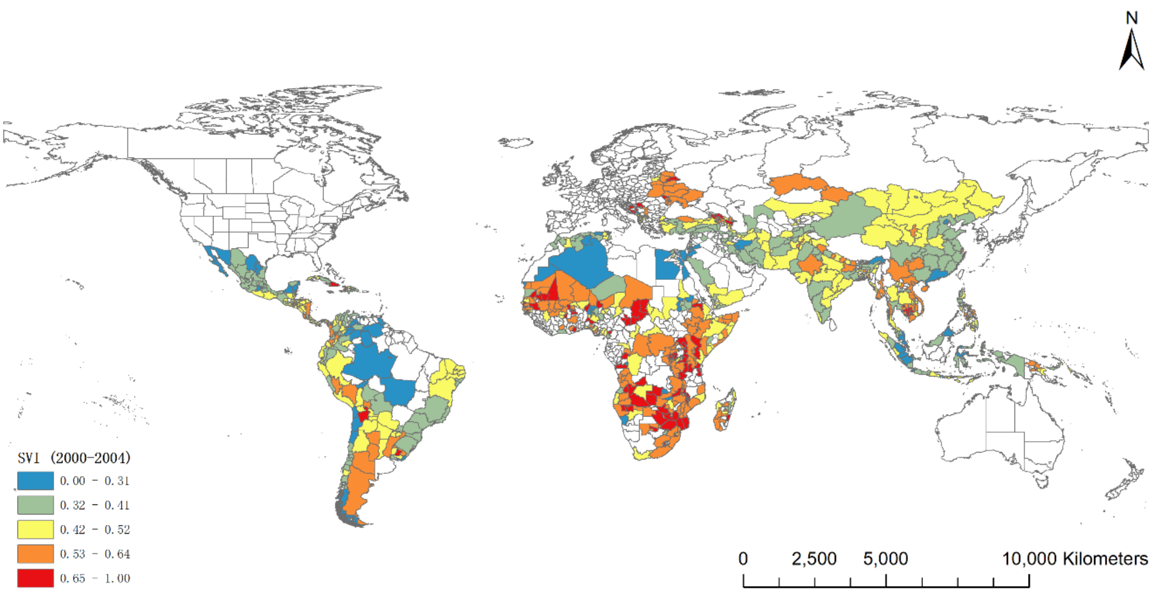

Figure 4. The global distribution of the social vulnerability index in period 2000-2004

Note: Orange (4) represents a subnational region with high-medium social vulnerability; Yellow (3) represents a subnational region with medium social vulnerability; Green (2) represents a subnational region with medium-low social vulnerability level; Blue (1) represents a subnational region with low social vulnerability.

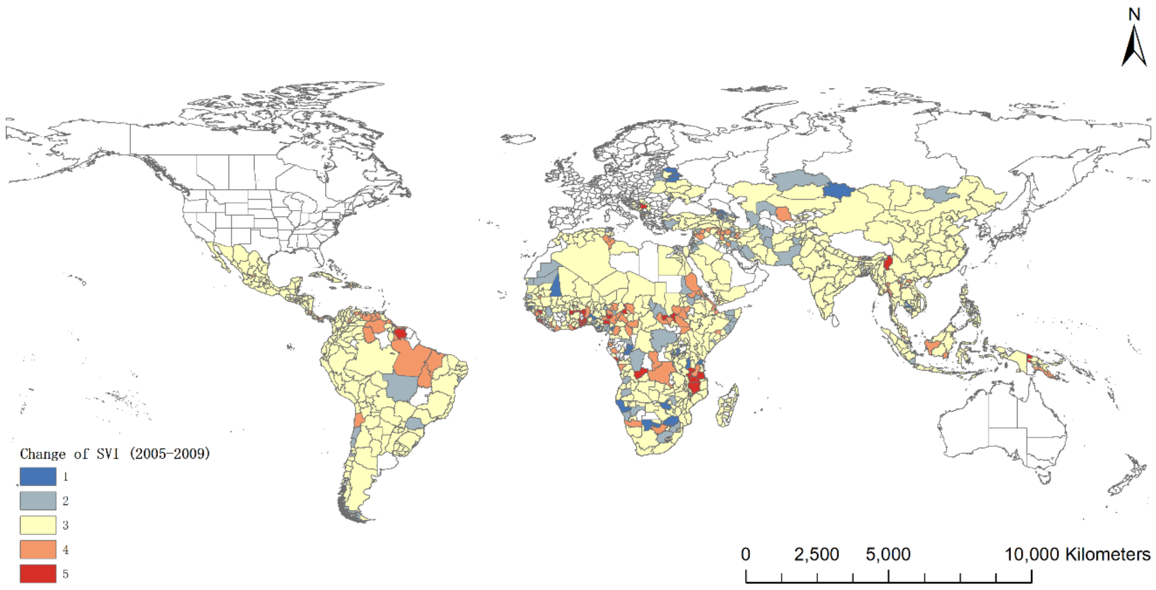

Figure 5. The change of the social vulnerability index in period 2005-2009 (compared with period 2000-2004)

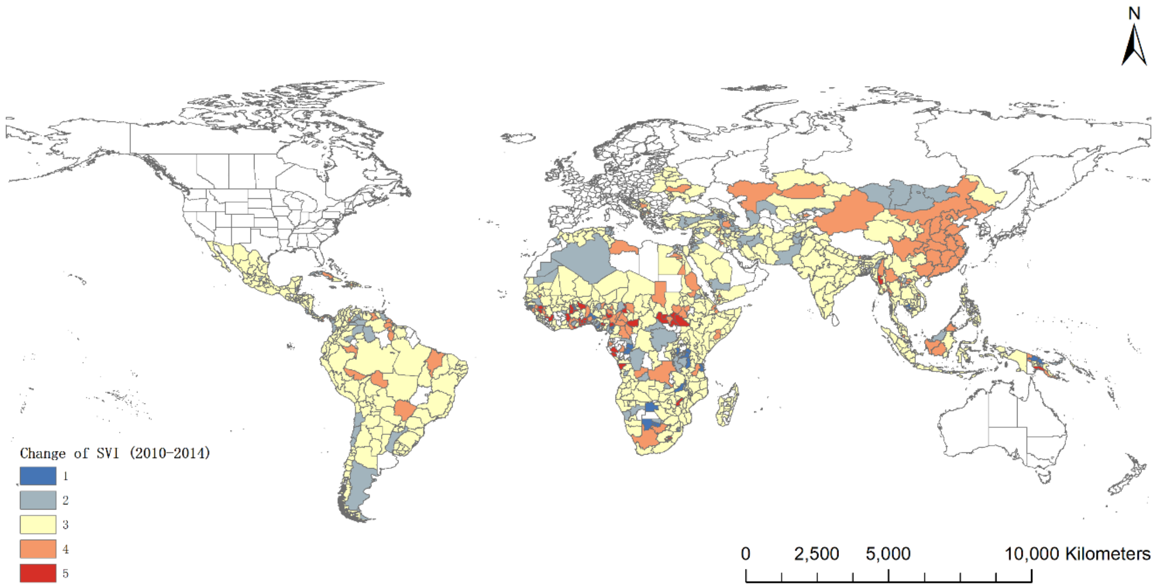

Figure 6. The change of the social vulnerability index in period 2010-2014 (compared with period 2000-2004)

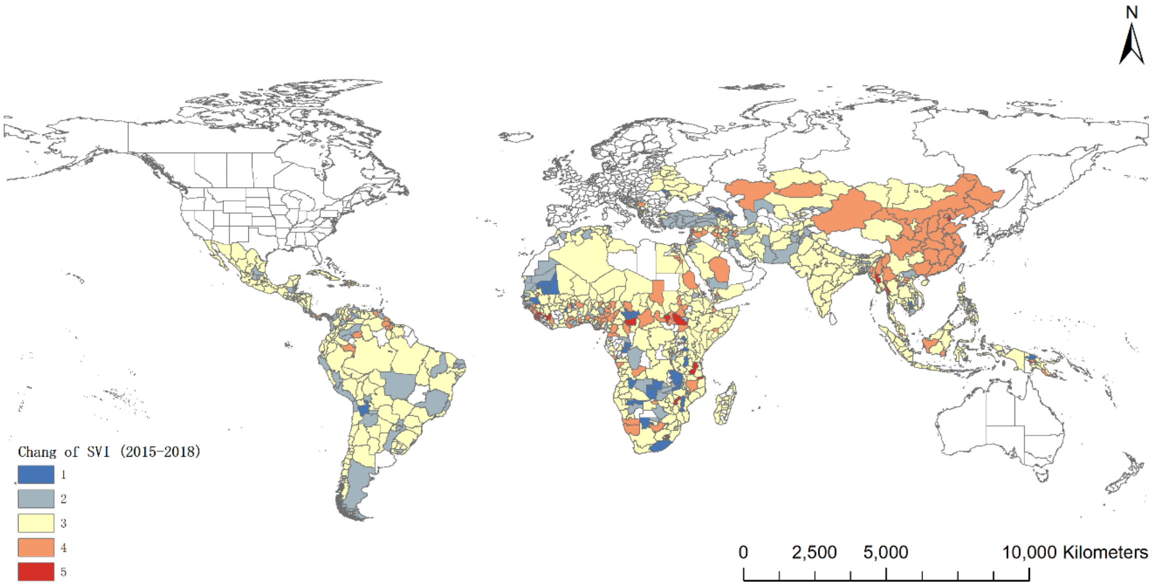

Figure 7. The change of the social vulnerability index in period 2015-2018 (compared with period 2000-2004)

Distinctive spatial distribution patterns of global social vulnerability at the subnational level are found from 2000 to 2018. The SVI score is standardized to the range of 0 to 1, indicating the level of social vulnerability increases from the lowest to the highest. Thus, the social vulnerability level of one subnational region is determined by comparison among all subnational regions in the studied period. The SVI in the first research period 2000-2004 is set as baseline to compare the change of social vulnerability in the next three periods. The change of the SVI score is divided into five categories by equal interval, indicating social vulnerability from decreasing to increasing from 1 to 5, while category 3 indicates almost no change between the referred period and the baseline period.

Shown in Fig.4, the most socially vulnerable subnational regions are mainly distributed in middle and southeast Africa, southern South America and western and southern Asia. Compared with the baseline period, northern South America, middle Africa and several subnational regions in western and southern Asia became more socially vulnerable to natural hazards. The level of social vulnerability of subnational regions in Africa still increased in the period of 2010-2014, while East Asia became increasingly vulnerable, especially in China. The pattern remained similar in the last period, this may due to the rapid aging of society in this time. The least vulnerable subnational regions locate mostly in North Africa and subnational region in the southeast Asia. Compared with the other regions at the same period, the social vulnerability of southern America and African sub-national regions slightly changed, while East Asia especially become more socially vulnerable to natural hazards.

5. Discussion and conclusion

5.1 Main findings

Our analysis of global social vulnerability to natural hazards reveals large spatial and temporal variation across subnational regions. To construct the Social Vulnerability Index, a set of demographic and socio-economic indicators identified in the previous literature is included. The result of principle component analysis shows that socio-economic development and population structure are the main components explaining most of the variance in the set of indicators. The SVI score was calculated by applying an inductive model with assumptions of the underlying vulnerability the principle component represents. The regions with high SVI score were mostly located in middle and southern Africa, west and southern Asia. Especially when compared with the baseline period, subnational regions in eastern Asia became more socially vulnerable as time passes. This may be due to a significant increase in the proportion of older populations in the region.

Corresponding to the previous studies, the economic status, education level both have significant influence on regional social vulnerability. Advanced economic growth and more educated population all lead to relatively low SVI. After incorporating the actual impact of disasters in the Poisson regression model, the results also indicate a negative relationship between economic status and education and disaster fatality. It also shows that life expectancy has significant negative effect on disaster fatality, while old age dependency is positively associated with disaster fatality. This may imply the intermediate effect of education or economic status on life expectancy of the elder population. Also, the model result suggested urbanization may increase the potential of adverse impact from disasters. High population density, inadequate and inefficient public facility and infrastructure in urban region may hinder the capacity of the region to cope and recover from extreme weather events.

Using subnational level data help capture more localized characteristics of the regions and help detect spatial pattern of social vulnerability which cannot be investigated by the previous studies at a national scale. In the context of climate change, it is crucial to understand the interior mechanism of social vulnerability through a validation with the actual impact of disasters. It is important to make appropriate institutional arrangements and improve standard of living for socially vulnerable population can help ease the risks they are facing.

5.2 Limitations and prospects

There always has been a disagreement concerning the selection of variables that represent and measure social vulnerability. Due to data availability and quality at the subnational level, the variables selected to construct SVI in this study may not be the most accurate or broad enough indicators to measure social vulnerability to natural disasters. Public facility and infrastructure, lifeline support and emergency management and risk awareness are beyond the scope of this study. Also, because the socio-economic data are not available at the same period, the imputed dataset is developed based on a linear trend assumption. In the next steps, we will look into the interaction among the socio-economic variables and adopt sensitivity analysis to test the constructed SVI. We will then construct a second set of social vulnerability index based on the model results. We also plan to collect more socio-economic data for the developed countries (which are not available in the Global Data Lab database) from Euro-stats, OECD regional statistics and other sources.

5.3 Conclusion

Based on the analysis of social vulnerability on the subnational level, this study finds distinctive and rather stable spatial patterns of SVI across different time periods, with the exception of eastern Asia. In general, subnational regions in Africa and South America are more socially vulnerable to natural hazards, but the social vulnerability of east Asia increases rapidly. Demographic and socio-economic factors both play important roles in determining the patterns of spatial distribution of SVI over the period observed. Economic development and population structure lead to the change of underlying driving forces of social vulnerability. Further investigation in identifying socio-economic clusters is required, in order to look into the different effect of demographic and socio-economic context on the actual impact of natural hazards.

References

Adger, & Kelly, M. (1999). Social Vulnerability to Climate Change and the Architecture of Entitlements. *Mitigation and Adaptation Strategies for Global Change*, *4*(3/4), 253–266. https://doi.org/10.1023/A:1009601904210

Adger, W. N. (1999). Social vulnerability to climate change and extremes in coastal Vietnam. *27*, *2*, 249–269.

Burton, C., Rufat, S [Samuel.], & Tate, E. (2018). Social Vulnerability: Conceptual Foundations and Geospatial Modeling.

CRED, & UNDRR. (2020). *Human cost of disasters: An overview of the last 20 years 2000-2019.*

Cutter, S. (1996). Vulnerability to environmental hazards. *Progress in Human Geography*, *20*(4), 529–539. https://doi.org/10.1177/030913259602000407

Cutter, S., Boruff, B., & Shirley, L. (2003). Social Vulnerability to Environmental Hazards. *Social Science Quarterly*, *84*(2), 242–261. https://doi.org/10.1111/1540-6237.8402002

Davino, C., Gherghi, M., Sorana, S., & Vistocco, D. (2021). Measuring Social Vulnerability in an Urban Space Through Multivariate Methods and Models. *Social Indicators Research*, *157*(3), 1179–1201. https://doi.org/10.1007/s11205-021-02680-0

Eakin, H., & Luers, A. L. (2006). Assessing the Vulnerability of Social-Environmental Systems. Annual Review of Environment and Resources, *31*, 365–394.

Fekete, A [A.] (2009). Validation of a social vulnerability index in context to river-floods in Germany. *Natural Hazards and Earth System Sciences*, *9*(2), 393–403. https://doi.org/10.5194/nhess-9-393-2009

Fekete, A [Alexander]. *Assessment of Social Vulnerability for River-Floods in Germany* [, Universitäts- und Landesbibliothek Bonn]. bonndoc.ulb.uni-bonn.de. https://bonndoc.ulb.uni-bonn.de/xmlui/handle/20.500.11811/4190

Flanagan, B. E., Hallisey, E. J., Adams, E., & Lavery, A. (2018). Measuring Community Vulnerability to Natural and Anthropogenic Hazards: The Centers for Disease Control and Prevention's Social Vulnerability Index. *Journal of Environmental Health*, *80*(10), 34–36.

Flanagan, B. E., Hallisy, E., Adams, E., & Lavery, A. (2018). Measuring community vulnerability to natural and anthropogenic hazards: The Center for Disease Control and Prevention's Social Vulnerability Index, *80*(10), 34–36.

Flatø, M., Muttarak, R [Raya], & Pelser, A. (2017). Women, Weather, and Woes: The Triangular Dynamics of Female-Headed Households, Economic Vulnerability, and Climate Variability in South Africa. *World Development*, *90*, 41–62. https://doi.org/10.1016/j.worlddev.2016.08.015

Fothergill, A. (1996). Gender, risk and disaster. *14*, *11*(33-56).

Frigerio, I., & Amicis, M. de (2016). Mapping social vulnerability to natural hazards in Italy: A suitable tool for risk mitigation strategies. *Environmental Science & Policy*, *63*, 187–196. https://doi.org/10.1016/j.envsci.2016.06.001

Garschagen, M., & Romero-Lankao, P. (2015). Exploring the relationships between urbanization trends and climate change vulnerability(133), 37–52.

Hoffmann, R., & Muttarak, R [Raya] (2017). Learn from the Past, Prepare for the Future: Impacts of Education and Experience on Disaster Preparedness in the Philippines and Thailand. *World Development*, *96*, 32–51. https://doi.org/10.1016/j.worlddev.2017.02.016

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A Program for Missing Data. *Journal of Statistical Software*, *45*(7). https://doi.org/10.18637/jss.v045.i07

Ignacio, J. A. F., Cruz, G. T., Nardi, F., & Henry, S. (2016). Assessing the effectiveness of a social vulnerability index in predicting heterogeneity in the impacts of natural hazards: Case study of the Tropical Storm Washi flood in the Philippines. *Vienna Yearbook of Population Research*, 91–129. https://doi.org/10.1553/populationyearbook2015s091

IPCC. (2012). *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*. Cambridge University Presss, Cambridge, UK,and New York, NY, USA, 582 pp.

Katharine, V. (2004). *Creating an Index of Social Vulnerability to Climate Change in Africa*. https://www.researchgate.net/profile/katharine-vincent-3/publication/228809913_creating_an_index_of_social_vulnerability_to_climate_change_in_africa

Kuhlicke, C., Scolobig, A., Tapsell, S., Steinführer, A., & Marchi, B. de (2011). Contextualizing social vulnerability: findings from case studies across Europe. *Natural Hazards*, *58*(2), 789–810. https://doi.org/10.1007/s11069-011-9751-6

Lutz, W., Muttarak, R [R.], & Striessnig, E. (2014). Universal education is key to enhanced climate adaptation. *346*, *6213*, 1061–1062.

Ma, L. (2016). Female labour force participation and second birth rates in South Korea. *Journal of Population Research*, *33*(2), 173–195. https://doi.org/10.1007/s12546-016-9166-z

Maiti, S., Jha, S. K., Garai, S., Nag, A., Chakravarty, R., Kadian, K. S., Chandel, B. S., Datta, K. K., & Upadhyay, R. C. (2015). Assessment of social vulnerability to climate change in the eastern coast of India. *Climatic Change*, *131*(2), 287–306. https://doi.org/10.1007/s10584-015-1379-1

Mallick, B., Rubayet Rahaman, K., & Vogt, J. (2011). Social vulnerability analysis for sustainable disaster mitigation planning in coastal Bangladesh. *Disaster Prevention and Management: An International Journal*, *20*(3), 220–237. https://doi.org/10.1108/09653561111141682

O'Neill, B., Jiang, L., KC, S., Fuchs, R., Pachauri, S., Laidlaw, E., Zhang, T., Zhou, W., & Ren, X. (2020). The effect of education on determinants of climate change risks. *3*, 520–528.

Otto, I. M., Reckien, D., Reyer, C. P. O., Marcus, R., Le Masson, V., Jones, L., Norton, A., & Serdeczny, O. (2017). Social vulnerability to climate change: A review of concepts and evidence. *Regional Environmental Change*, *17*(6), 1651–1662. https://doi.org/10.1007/s10113-017-1105-9

Reisinger, Andy, Mark Howden, & Carolina Vera. (2020). *The Concept of Risk in the IPCC Sixth Assessment Report: A Summary of Cross-Working Group Discussions.* Geneva, Swizerland. pp15. Intergovernmental Panel on Climate Change.

Rodenbiker, J. (2020). Urban Ecological Enclosures: Conservation Planning, Peri-urban Displacement, and Local State Formations in China, *44*(4), 691–710.

Rosvold, E. L., & Buhaug, H. (2021). Gdis, a global dataset of geocoded disaster locations. *Scientific Data*, *8*(1), 61. https://doi.org/10.1038/s41597-021-00846-6

Rufat, S [Samuel], Tate, E., Burton, C. G., & Maroof, A. S. (2015). Social vulnerability to floods: Review of case studies and implications for measurement. *International Journal of Disaster Risk Reduction*, *14*, 470–486. https://doi.org/10.1016/j.ijdrr.2015.09.013

Rufat, S [Samuel], Tate, E., Burton, C., & Maroofb, A. (2015). Social vulnerability to floods: Review of case studies and implications for measurement, *14*(4), 470–486.

Rufat, S [Samuel], Tate, E., Emrich, C. T., & Antolini, F. (2019). How Valid Are Social Vulnerability Models? *Annals of the American Association of Geographers*, *109*(4), 1131–1153. https://doi.org/10.1080/24694452.2018.1535887

Sapam, S., Mohammad, E., & Sarbjeet, S. (2014). The concept of social vulnerability: A review from disasters perspectives, *1*(6), 71–82. http://www.drbrambedkarcollege.ac.in/sites/default/files/vulnarability.pdf

Siagian, T. H., Purhadi, P., Suhartono, S., & Ritonga, H. (2014). Social vulnerability to natural hazards in Indonesia: Driving factors and policy implications. *Natural Hazards*, *70*(2), 1603–1617. https://doi.org/10.1007/s11069-013-0888-3

Smits, J., & Permanyer, I. (2019). The Subnational Human Development Database. *Scientific Data*, *6*, 19–38. https://doi.org/10.1038/sdata.2019.38

Tate, E. (2012). Social vulnerability indices: A comparative assessment using uncertainty and sensitivity analysis. *Natural Hazards*, *63*(2), 325–347. https://doi.org/10.1007/s11069-012-0152-2

United Nations. (2015). *Transforming our world: the 2030 Agenda for Sustainable Development.*

UNU-EHS. (2008). *Megacities: urbanization and social vulnerability*. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.173.8169&rep=rep1&type=pdf#page=22

Wisner, B., Blaikie, P., Cannon, T., & Davis, I. (2003). *At Risk: Natural hazards, people's vulnerability and disasters*. Second edition.

Wolkin, A., Patterson, J. R., Harris, S., Soler, E., Burrer, S., McGeehin, M., & Greene, S. (2015). Reducing Public Health Risk During Disasters: Identifying Social Vulnerabilities. *Journal of Homeland Security and Emergency Management*, *12*(4), 809–822. https://doi.org/10.1515/jhsem-2014-0104

Wood, E., Sanders, M., & Frazier, T. (2021). The practical use of social vulnerability indicators in disaster management. *International Journal of Disaster Risk Reduction*, *63*, 102464. https://doi.org/10.1016/j.ijdrr.2021.102464

Zhou, Y., Li, N., Wu, W., Wu, J., & Shi, P. (2014). Local spatial and temporal factors influencing population and societal vulnerability to natural disasters. *Risk Analysis : An Official Publication of the Society for Risk Analysis*, *34*(4), 614–639. https://doi.org/10.1111/risa.12193

EMDAT. EM-DAT, the International Disaster Database. 2022. https://public.emdat.be/data (accessed 16 May 2022)