

RESEARCH

Open Access

A statistical analysis of the global historical volcanic fatalities record

Melanie Rose Auker^{1*}, Robert Stephen John Sparks¹, Lee Siebert², Helen Sian Crosweller¹ and John Ewert³

Abstract

A new database of volcanic fatalities is presented and analysed, covering the period 1600 to 2010 AD. Data are from four sources: the Smithsonian Institution, Witham (2005), CRED EM-DAT and Munich RE. The data were combined and formatted, with a weighted average fatality figure used where more than one source reports an event; the former two databases were weighted twice as strongly as the latter two. More fatal incidents are contained within our database than similar previous works; approximately 46% of the fatal incidents are listed in only one of the four sources, and fewer than 10% are in all four. 278,880 fatalities are recorded in the database, resultant from 533 fatal incidents. The fatality count is dominated by a handful of disasters, though the majority of fatal incidents have caused fewer than ten fatalities. Number and empirical probability of fatalities are broadly correlated with VEI, but are more strongly influenced by population density around volcanoes and the occurrence and extent of lahars (mudflows) and pyroclastic density currents, which have caused 50% of fatalities. Indonesia, the Philippines, and the West Indies dominate the spatial distribution of fatalities, and there is some negative correlation between regional development and number of fatalities. With the largest disasters removed, over 90% of fatalities occurred between 5 km and 30 km from volcanoes, though the most devastating eruptions impacted far beyond these distances. A new measure, the Volcano Fatality Index, is defined to explore temporal changes in societal vulnerability to volcanic hazards. The measure incorporates population growth and recording improvements with the fatality data, and shows *prima facie* evidence that vulnerability to volcanic hazards has fallen during the last two centuries. Results and interpretations are limited in scope by the underlying fatalities data, which are affected by under-recording, uncertainty, and bias. Attempts have been made to estimate the extent of these issues, and to remove their effects where possible.

The data analysed here are provided as supplementary material. An updated version of the Smithsonian fatality database fully integrated with this database will be publicly available in the near future and subsequently incorporate new data.

Keywords: Volcanoes, Fatalities, Population, Volcano Fatality Index

Introduction

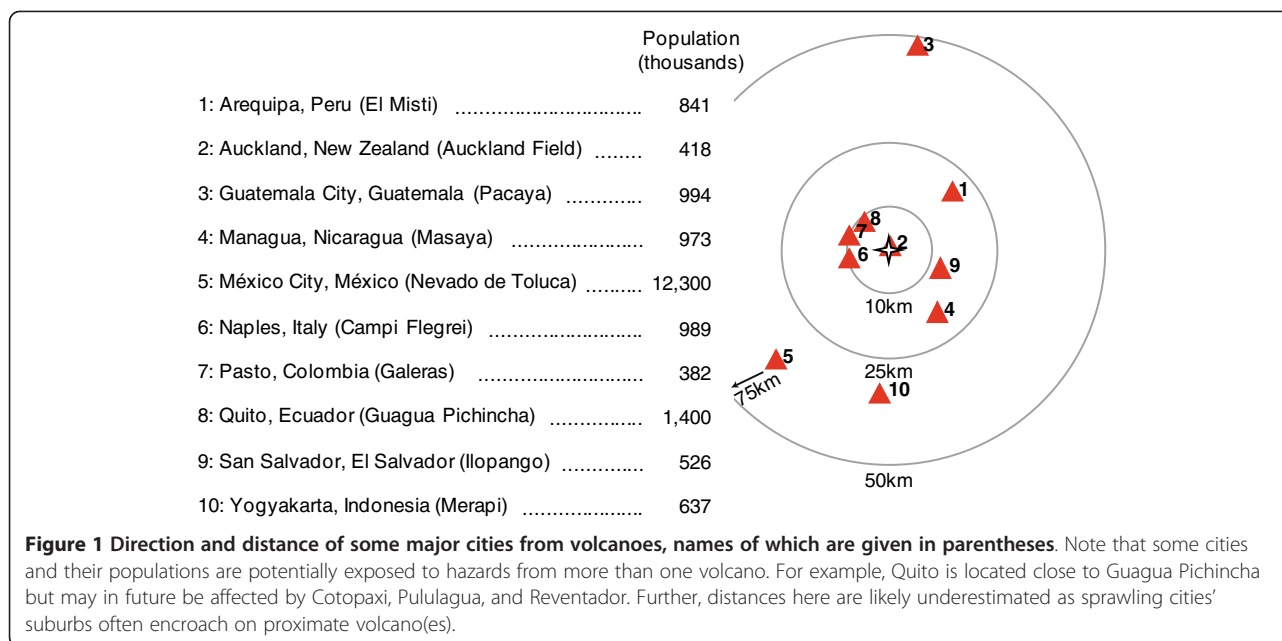
The advent of the 21st century has seen society's exposure to volcanic hazards rapidly increase as a result of population growth and development (Chester et al., 2001). Expansion of urban centres located close to active volcanoes is increasing the number of people at risk (Figure 1). By updating the analysis of Peterson (1986) and Small and Naumann (2001) using 2009 World Bank population data, it is estimated that at least 600 million people live in areas potentially affected by volcanic hazards. Developments in

volcanology, such as improved monitoring techniques, better understanding of eruption precursors, and more confident forecasting (Sparks et al., 2012), along with more proactive mitigation measures (predominantly evacuation), should reduce vulnerability and counteract increased exposure of societies to volcanic hazards (e.g. Donovan et al., 2012).

This paper investigates the historical record of human mortality from volcanic activity, with the aim of giving insights into the causes of volcanic disasters, and understanding the main controls on loss of life and how the volcanic fatalities record has changed over time. Questions we seek to answer include the following: how are volcanic

* Correspondence: ma6549@bristol.ac.uk

¹Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK
Full list of author information is available at the end of the article



fatalities distributed in space and time? Which volcanic hazards have caused most fatalities? To what extent are the data affected by under-recording and biases, and how can these be taken into account? What do the data reveal are the main controls on vulnerability to volcanic disasters and how have these changed with time due to population growth, urbanisation, and advances in modern science?

Data sources

Various databases covering volcanic fatalities, injuries, evacuations, economic damages, and other variables have been compiled. Whilst the quality of those that are maintained by scientific organisations using volcanological bulletins and peer-reviewed literature is high, others that use news and non-governmental agency reports as their main sources are less reliable, though useful nonetheless (Guha-Sapir and Below, 2006). A third type of resource has also emerged, aiming to compile and verify the aforementioned databases; examples include Witham (2005) and Tanguy et al. (1998). We do not use the Tanguy et al. (1998) database as a source here as it covers only the largest disasters.

Databases

This study uses four main sources of data, and consequently contains a greater number of fatal incident records than other works before it. The Smithsonian Institution's (SI) Global Volcanism Program compiles records of worldwide Holocene volcanoes and their eruptions, including a database of fatal volcanic incidents. The version of the SI database used here covers the period 4350 BC to 2008, and holds the number of fatalities, split of fatalities across causes (such as pyroclastic density

currents (PDCs) and mudflows (lahars), see Table 1), information on the distance of fatalities from the volcano, and reference(s) for 565 events. Events are included in the SI database if they have caused one or more fatality. The SI fatalities database gathers the majority of its information from volcano observatories, scientific literature, and volcanological bulletins from verifiable sources, though earlier entries rely more heavily on non-scientific literature.

Witham (2005) presents "a new database on human mortality and morbidity, and civil evacuations arising from volcanic activity" for events in the 20th century (page 191). The compilation includes data for numbers of fatalities, injuries, evacuees, and people made homeless, as well as fatality cause(s), information on the distance of fatalities from the volcano, and reference(s). Events are included if they have affected humans; 491 are listed, of which roughly 53% have caused loss of life. Where possible, more than one source is used for each event; any discrepancies are noted and a "best" estimate recorded. Data are predominantly sourced from Volcanoes of the World (Siebert and Simkin, 2002; Siebert et al. 2010), SI Global Volcanism Program activity reports (Venzke et al., 2002-), the Catalogue of Active Volcanoes (IAVCEI, 1951–1975), and the Bulletin of Volcanic Eruptions (Volcanological Society of Japan, 1960–1996).

Munich Reinsurance (Munich RE) set up a research unit for natural hazards in 1970. Various datasets have been developed since, one of which covers volcanic events that have "either caused any property damage, or injured or killed any person". The database reports the number of fatalities, overall losses, and insured losses, as well as a written incident description, for 452 events occurring between 1950 and 2009 (data donated by Wirtz,

Table 1 List of fatality causes used in analyses

| Cause | Definition | Abbreviation |
|-------------------------------------|--|--------------|
| Pyroclastic density currents (PDCs) | Pyroclastic flows, surges, and directed blasts | P |
| Tephra | Tephra (ash, bombs, lapilli), resultant from both magmatic and phreatic eruptions. Death either by ballistic impact, or, with finer-grained ash, by suffocation, collapse of ash-covered roofs etc | T |
| Primary Lahars (Mudflows) | Lahars at time of eruption | M |
| Indirect | Indirect deaths (disease, starvation, exposure) | I |
| Waves (Tsunami) | Waves or tsunami | W |
| Secondary Lahars (Mudflows) | Post-eruption lahars | m |
| Lava Flows | Lava flows | L |
| Avalanches | Debris avalanches and landslides | A |
| Gas | Gas emission from eruptive craters, as well as fumarolic/ solfataric activity | G |
| Seismicity | Seismic, or volcanic earthquake (tectonic earthquake deaths excluded) | S |
| Lightning | Lightning strike (electrostatic discharge) | E |
| Floods (Jökulhlaups) | Floods and Jökulhlaups | F |

2009). Because of the industry in which Munich RE operates, the primary interest is monetary losses and many of the data are consequently derived from the insurance business, such as Munich RE branch offices, insurance associations, and insurance press. Some scientific sources, governments, weather services, press agencies, and NGOs are also used. Some of the data are not easily verified independently due to their commercial or grey literature origin, though a ranking system of data trustworthiness and quality is used to improve overall data quality (Wirtz, 2009, pers. comm.).

The Centre for Research on the Epidemiology of Disasters (CRED) started to compile a database of natural disasters, named EM-DAT, in 1990; it “serves the purposes of humanitarian action at national and international levels” (The Centre for Research on the Epidemiology of Disasters, 2010). The database holds 220 publicly-available records of volcanic disasters from 1900 to 2011, listing the number of people killed, the total number of people affected, and estimated damage in US dollars. An event is included in the database if it meets one or more of the following criteria:

- Declaration of a state of emergency or call for international assistance
- 10 or more people reported killed
- 100 or more people reported affected.

The first of these takes precedence for entry, even if neither of the latter two is fulfilled. Similarly to the Munich RE database, EM-DAT is less heavily based on scientific papers and reports, and instead uses more information from governments, aid agencies, and insurance companies.

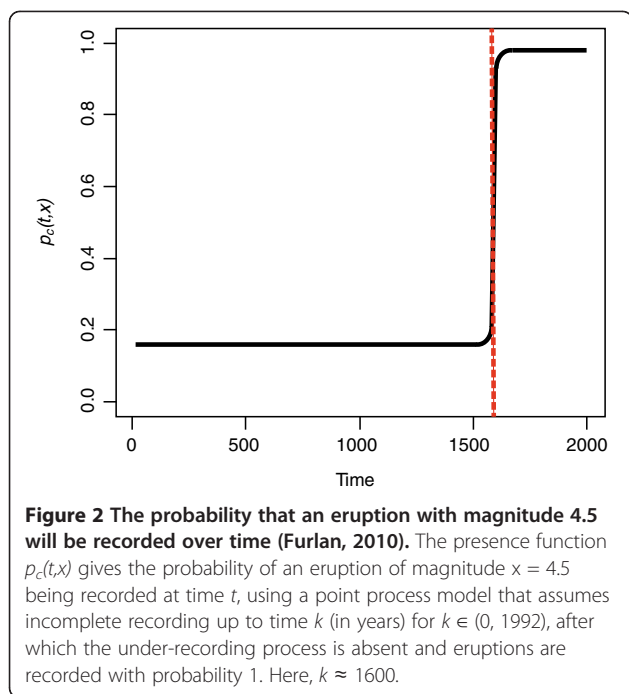
Data and database issues

In using four sources, we have created a well-referenced database with wide-ranging spatial and temporal coverage. However, there are known weaknesses that affect such databases that combining four sources cannot overcome.

Firstly, a balance must be sought between the number of data sources used and their rigour. On one hand, using only a few data sources known to be very well referenced from peer-reviewed scientific literature yields a more accurate final optimum dataset. On the other hand, inclusion of datasets that are less accurate increases the total amount of data available to use and number of volcanic incidents covered.

Secondly, there is a trade-off between the length of time covered by a data source and the availability of data; records are scarcer back in time. A study period that is long enough to contain sufficient records for statistically sound analysis, but not so long that the earliest entries are shrouded in uncertainty, should be chosen. Statistical modelling of historical magnitude-frequency data is useful in optimising temporal coverage. We use the analysis of Furlan (2010), who presents a change-point model of the censoring effect in the recording of volcanic eruptions of magnitude 4 or greater. The Furlan study shows a marked improvement in the probability of an eruption being recorded at the turn of the 17th century (Figure 2). Consequently, we only investigate eruptions occurring in or after 1600 AD.

Lastly, human impact variables are heterogeneous with respect to their availability and accuracy. Whilst almost all relevant source databases report the number of people killed by an eruption, some do not record economic



damages, or numbers injured, evacuated, and made homeless; there are no standardised definitions of these four impact variables (Guha-Sapir and Below, 2006). Of those databases that do record multiple impact facets, it is likely that some parameters are quoted far more accurately than others; the number of fatalities is widely regarded as the measure likely to be most accurately recorded. We thus analyse only volcanic fatalities data.

Methods

The four sources outlined above were compiled to produce our optimised database. We combine scientifically sound databases (SI and Witham, both of which incorporate data from Tanguy et al. (1998)) with more media-based sources (Munich RE, EM-DAT) to maximise the amount of reliable data available.

Each record was listed as a single row in our database, with up to four sets of data, corresponding to the four sources. A weighted average was then calculated to give a single preferred fatality figure used for numerical analysis. The methods used in compiling this database are outlined below.

Preliminary editing

We used volcano names and volcano numbers as listed in the SI Volcano Reference Files (www.volcano.si.edu) for consistency. The subregion, country, and region were included for each event for spatial analyses. These are the island or sea the volcano is located on or in, the country to which the subregion belongs, and one of 16 regions, respectively. We used regions based on those defined by

IAVCEI and adopted by the SI (Table 2), though with adaptations for clarity.

Records were only included once in our database, and only if the source volcano could be determined unambiguously. Repeated records were excluded, along with those where, even after further research, errors of multiple, missing, or unrecognised volcano names could not be corrected. A total of 33 records across the Witham, Munich RE, and EM-DAT databases were excluded as a result (Table 3). A further record was excluded from the Witham database because the fatalities resulted from non-volcanic causes.

The date (or date range, for some entries in the Munich RE database) is the date that an event caused human impacts. There were some date errors that became evident when compiling the database; two changes were made to data originating from each of the SI, Munich RE, and EM-DAT databases.

Prior to numerical analysis, the notation used for indirect fatalities in the SI database, and that for secondary and tertiary estimates of fatality figures, was modified. For two records, indirect fatalities were listed as a separate entry to fatalities caused directly. This is problematic for later analysis (each record must be displayed as a single row), and we thus combined the direct and indirect data into a single entry. For some large events (e.g. Tambora, 1815) the SI database gives multiple fatality counts to document varying estimates from different sources, with one identified as the preferred value. To avoid multiple counting, we chose the preferred fatality figure and associated database entry. A total of 9 non-preferred records were excluded.

Several methods of dealing with qualitative frequency words have been utilised in damage intensity scales (Blong, 2003). In this work, we follow the scheme of Simkin et al. (2001), listed in Table 4, to convert qualitative fatality entries to quantitative ones; this allows numerical investigation of the entire fatalities record. In cases where fatalities occurred but the number was unknown, a value of 15 was used, again in line with the figures used by Simkin et al. (2001). Such qualitative-to-quantitative formatting was required only for the SI database; fatality counts in the Munich RE and EM-DAT databases are solely numerical, and those in the Witham database are quantitative as the conversion of qualitative data has already been applied, also following Simkin et al. (2001).

Along with qualitative entries, the SI and Witham databases also include uncertain fatality figures denoted by questions marks, tildes, or similar; for example “~100” and “45?”. The approximate figure was taken as a quantitative best estimate value in these cases.

A total of 170 of the 554 fatality figures listed in the SI source database are quoted with some level of uncertainty,

Table 2 IAVCEI (1951 – 1975) volcanic region names, and adapted names as used here

| Region – IAVCEI name | Region – Name in this paper | Countries contained within region |
|---|-----------------------------|--|
| Mediterranean and West Asia | Mediterranean | Italy Greece Turkey |
| Africa and Red Sea | Africa and Red Sea | Cameroon Democratic Republic of Congo Ethiopia Red Sea (off-shore Yemen) Tanzania |
| Middle East and Indian Ocean | Indian Ocean | Comoros Reunion |
| New Zealand to Fiji | New Zealand to Fiji | New Zealand Tonga |
| Melanesia and Australia | Melanesia | Papua New Guinea Solomon Islands Vanuatu |
| Indonesia | Indonesia | Indonesia |
| Philippines and South East Asia | Philippines and SE China | SE China Philippines |
| Japan, Taiwan, Marianas | Japan | Japan |
| Kuril Islands + Kamchatka and Mainland Asia | Kuril Islands and Kamchatka | Kuril Islands Kamchatka |
| Alaska + Canada and Western USA | North America | Aleutian Islands Alaska Canada USA (excluding Hawaii) |
| Hawaii and Pacific Ocean | Hawaii | Hawaii |
| Mexico and Central America | Mexico and Central America | Costa Rica El Salvador Guatemala México Nicaragua |
| South America | South America | Chile Colombia Ecuador Perú |
| West Indies | West Indies | Martinique Montserrat St. Vincent and the Grenadines |
| Iceland and Arctic Ocean | Iceland | Iceland |
| Atlantic Ocean | Atlantic Ocean | Azores (Portugal) Canary Islands (Spain) Cape Verde |

Note that the countries listed in Table 2 are only the 38 in which fatalities have been recorded; the IAVCEI regional definitions include additional countries, but they do not have fatalities recorded in our database and are thus excluded from analyses. The names used in this paper are designed to reflect this focus on fatalities.

Table 3 Records excluded from the source databases for reasons related to volcano names

| Database | Reason for exclusion | Number of entries excluded |
|-----------|----------------------|--|
| Witham | Name unrecognised | 1 |
| Munich RE | Name ambiguous | 2 |
| | Name unknown | 7 |
| | Name unrecognised | 6 |
| | Doubled record | 6 |
| EM-DAT | Name unknown | 2 |
| | Name unrecognised | 1 |
| | Doubled record | 6 |
| | Trebled record | 2 (1 event, 2 superfluous records removed) |

with a corresponding 80 of 536 in the Witham database; due to overlap in the SI and Witham databases, our combined database contains 209 uncertain fatality counts (Table 5).

Our database contains 1,115 individual volcanic events spanning 4350 BC to 2011; 1060 occurred after 1600 AD (our period of study here). 694 events are listed in only one of the four sources, 274 are listed in two sources, 100 are listed in three sources, and 47 are in all four. The SI database contributes 554 records to our database, the Witham database contributes 536, and the Munich RE and EM-DAT sources contribute 431 and 209 records respectively. As discussed later, these data are incomplete as low-magnitude, low-fatality eruptions in earlier centuries were likely under-reported.

Creating an incidents database

After the formatting outlined above, records were streamlined to create an incidents database, containing only those events which have caused human impacts of any kind. Here we should make our nomenclature clear. We follow the SI approach for the term “event”, using it to refer to any volcanic episode (non-eruptive and secondary phenomenon, as well as eruptions) recorded in the four source databases, irrespective of whether it caused any human impacts. We use the term “incident” to identify those events that caused human impacts of any kind, such

Table 4 Enumerations used to change qualitative descriptions of fatality counts to quantitative, to allow numerical analyses

| Qualitative description | Quantitative value, from Simkin et al. (2001) |
|-------------------------|---|
| Few | 3 |
| Some | 3 |
| Several | 5 |
| Unknown | 15 |
| Many | 100 |
| Hundreds | 300 |

as fatalities, injuries, evacuations, or homelessness. Whilst in theory all four sources only list incidents, in practice there are inconsistencies in the Munich RE and EM-DAT databases where events are recorded despite not meeting the database’s entry criteria. Definitions used for the removal of events not classed as incidents are:

- Munich RE: the number of fatalities, insured losses, and uninsured losses are all zero; the written description details no damage, evacuations, people made homeless, or people affected; no other database has a record of the event that contains impacts of any kind.
- EM-DAT: the number of fatalities, total affected, and damage are all zero; no other database has a record of the event that contains impacts of any kind.

Following application of these formatting rules, 95 event records were excluded from the Munich RE database, and 9 from the EM-DAT. One event was common to both sources.

The incidents database contains 1,012 individual volcanic incidents spanning 4350 BC to 2011; 957 occurred after 1600 AD (our period of study here). 592 incidents are listed in only one of the four sources, 273 are listed in two sources, 100 are listed in three sources, and 47 are in all four. The SI database contributes 554 records to the incidents database, the Witham database contributes 536, and the Munich RE and EM-DAT sources contribute 336 and 200 records, respectively.

Creating a fatalities database

We move from the incidents to the fatalities database by excluding incidents from the former for which the sum of fatalities across all four source databases is zero, i.e. no source database lists any fatalities. 249 records were excluded from the Witham database, 243 were excluded from the Munich RE database, and 107 were excluded from the EM-DAT. Note the term “fatal incident” is used to refer to records in our fatalities database. Unless

Table 5 Number of fatality figures in the SI and Witham databases which are given with some degree of uncertainty

| Number | Uncertainty notation | Number of such entries in SI database, post-1600 | Number of such entries in Witham database, post-1600 |
|--------|---|--|--|
| 1 | Completely unknown (“?”) | 48 | 7 |
| 2 | “Few” or “Few?” | 3 | 2 |
| 3 | “Some” or “Some?” | 4 | 3 |
| 4 | “Several” or “Several?” | 8 | 0 |
| 5 | “Many” or “Many?” | 21 | 1 |
| 6 | “Hundreds” | 1 | 0 |
| 7 | Approximate (“~x”) | 9 | 2 |
| 8 | Uncertain (“x?”) | 57 | 6 |
| 9 | Below a certain threshold (“<x”) | 1 | 0 |
| 10 | Above a certain threshold (“>x”) | 13 | 3 |
| 11 | Above an uncertain threshold (“>x?”) | 5 | 1 |
| 12 | Two options(“x or y”) | 0 | 2 |
| 13 | Two options, with one uncertain(“x or y?”) | 0 | 2 |
| 14 | Range with certain endpoints (“x-y”) | 0 | 42 |
| 15 | Range with one uncertain endpoint (“x - >y” or “> x - y”) | 0 | 9 |
| | Total | 170 | 80 |

Number of fatal incidents with uncertain fatality figures: 209.

otherwise stated, where we refer to the largest disasters we are concerned with those which have caused the greatest number of fatalities.

The SI database contributes 554 records to the fatalities database, the Witham database contributes 287, and the Munich RE and EM-DAT sources each contribute 93 records. A total of 588 individual fatal incidents occurring between 4350 BC and 2010 are detailed; 533 of these occurred between 1600 and 2010, and it is these fatal incidents we use for numerical analysis. 244 of the post-1600 fatal incidents are listed in only one of the four sources, 186 are listed in two sources, 56 are listed in three sources, and 47 are in all four.

For the 289 fatal incidents for which two or more sources list fatalities, some process must be employed to combine fatality counts into one representative figure. We used a weighted arithmetic average, with weightings guided by the sources, methods, and reputation of the four databases. Whilst none of the four source databases are perfect, the formatting undertaken to arrive at the fatalities database has highlighted far more weaknesses in the Munich RE and EM-DAT databases than in those of the SI and Witham. The former two are based more heavily on non-scientific ancillary descriptions, whilst the latter two reference peer-reviewed literature and volcanological bulletins more heavily. Based on these reasons, we weighted fatality counts in the SI and Witham databases twice as strongly as those in the Munich RE and EM-DAT databases; the weighted figures were then averaged to give “best estimate” fatality figures, used in

all further analyses. Note that as a result of the averaging process, there are two fatal incidents for which the “best estimate” fatality figure is 0.25, rounded down to zero for all analyses.

For 96 of the 289 fatal incidents for which two or more sources record a fatal incident, there is disagreement regarding the number of fatalities. Such disagreements often arise as a result of multiple sources reporting fatality figures, such as aid agencies, media outlets, and scientists; the former two of these groups may have incentives to exaggerate the extent of disaster in order to secure more aid or make more sensational news, respectively (Blong, 1984). Fatality counts over which there are discrepancies can be used as a simple measure of data uncertainty, providing insight into the extent of these misreporting issues; for each fatal incident, the average of each of the fatality counts is calculated and the percentage departure from this average evaluated for each source’s fatality estimate. A histogram of the percentage departures is shown in Figure 3 and is approximately normally distributed. The standard deviation around the mean is 51% and almost all the percentage departure values are between 30% and 300%. This suggests that the uncertainties in fatality counts vary by typically no more than a factor of three, and most vary by less than a factor of two. Whilst this method does not fully capture all the sources of uncertainty, it is a good indicator of the non-systematic contribution to the uncertainty in fatality figures. We note that in 193 cases where an incident is listed in more than one database, there is agreement about the number of fatalities.

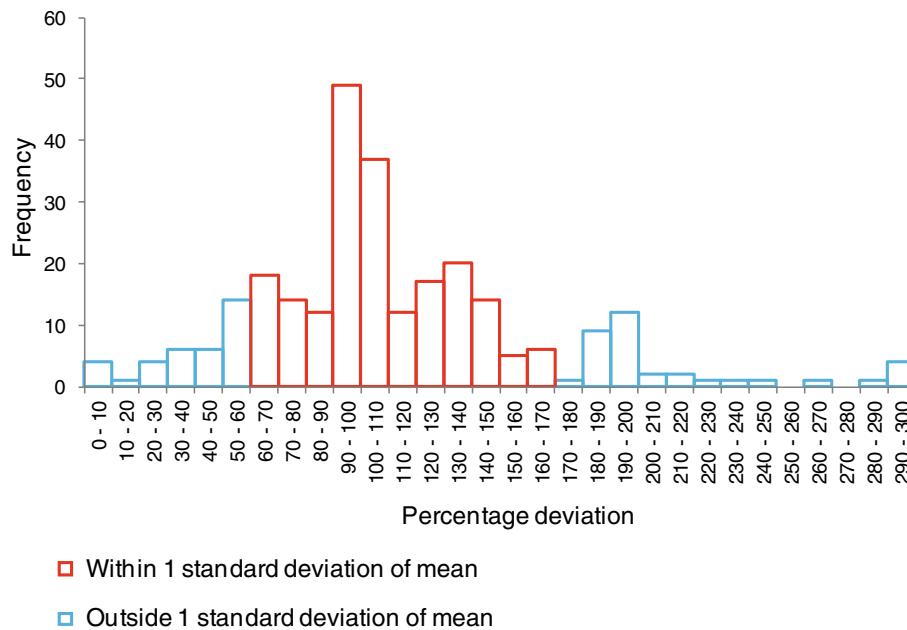


Figure 3 Percentage deviation around mean average fatality figure for cases where more than one database lists a fatality figure (96 fatal incidents).

This is likely attributed to each database obtaining the information from the same source; the Witham, Munich RE and EM-DAT databases often use data from the SI.

Supplementary datasets

Along with the data available from our fatalities database, we use additional datasets covering Human Development Index (HDI) and population. HDI data are from the United Nations Development Program (UNDP, 2011), and cover the period 1980 to 2010 for 33 of the 38 countries in the database. HDI figures were unavailable for Reunion, St. Vincent and the Grenadines, Montserrat, Martinique, and Vanuatu. We thus use France's HDI figures as a proxy for Reunion, and the Caribbean region's for St. Vincent and the Grenadines, Montserrat, and Martinique (no data for Vanuatu). HDI values were available for most countries for the years 1980, 1990, 2000, 2005, and 2010. We calculate regional representative HDI figures by averaging over constituent countries (as listed in Table 2), and time. We follow the UN's classification method for development levels, using quartiles to divide the regions into low, medium, high, and very high development levels. Quartiles were derived from global HDI data, rather than just our subpopulation of countries with volcanic fatalities, to give a more comprehensive indication of the absolute development level of the 16 regions.

Annual population data were available for the period 1600 AD to 2010 from three sources: McEvedy and Jones (1978); Populstat (1996/ 2006); and the Population Division, Population Estimates, and Projections Section of the UN

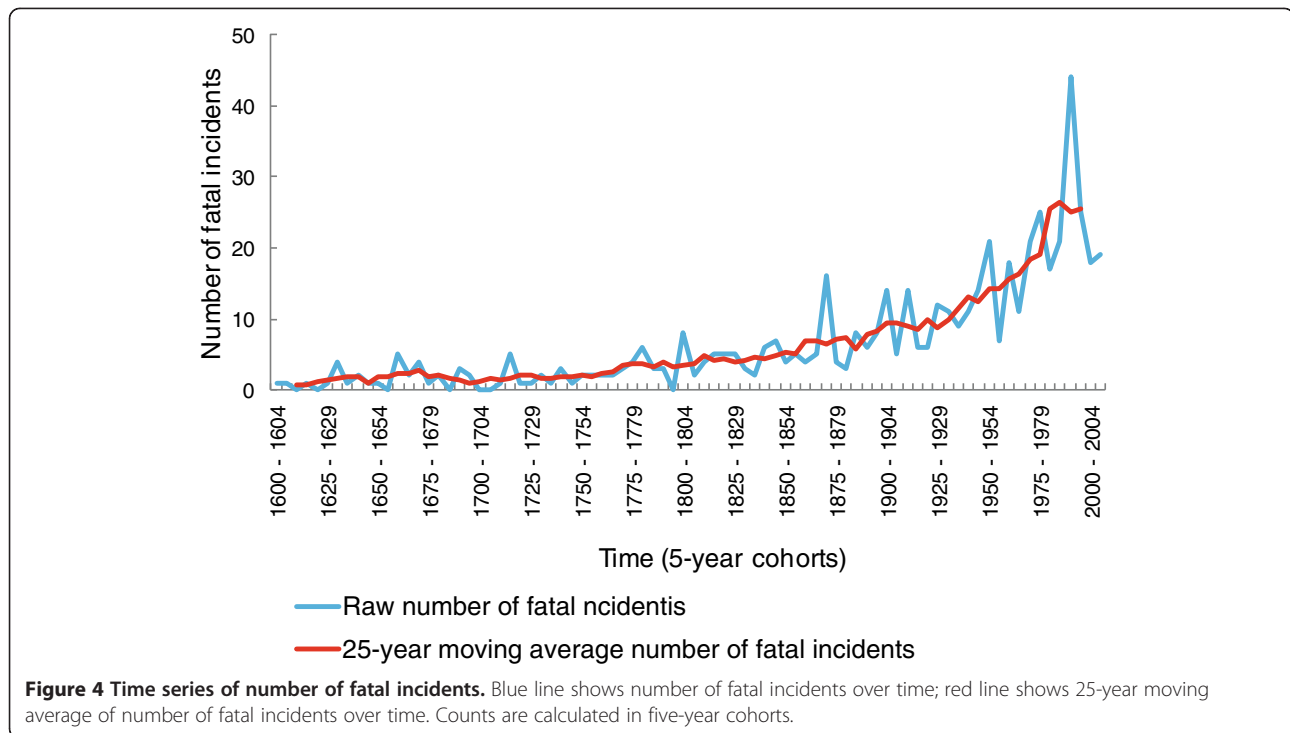
Department of Economic and Social Affairs (2010). After combining these three sources, many data points were still missing. Exponential growth models were fitted to each country's known population data, and these models were used to infer all missing population data points. This method provides an estimated annual time series of population data for 37 of the 38 countries in which volcanic fatalities have been recorded for the period 1600 AD to 2010 (no population data were available for Montserrat).

Results

The finalised fatalities database holds records of 588 fatal volcanic incidents occurring between 4350 BC and 2010. The details of 533 fatal volcanic incidents, which occurred between 1600 AD and 2010, were analysed. Our investigation includes consideration of the uncertainties, data completeness and data quality.

Time series analyses

Figure 4 shows the number of fatal incidents recorded in five-year cohorts from 1600 to 2010. Though the raw data are noisy, the general trend (25-year moving average) shows a fairly constant rate of recording through the 17th and 18th centuries, then an increasing rate to present day. The increasing number of fatal incidents occurring from roughly 1800 to 2010 is likely attributable to a combination of improved recording of all volcanic events (both fatal and non-fatal), along with a real increase in the number of fatal incidents.



In Figure 5, we attempt to disentangle the effects of improving recording and real increases in the number of fatal incidents. A total of 5,815 volcanic events have been recorded from 1600 to 2010 (using data from the SI); an average 9.2% have caused reported fatalities. As a time series of 25-year moving average figures using five-year data cohorts, the percentage of all events causing fatalities varies between approximately 5 and 16%. There is much fluctuation but no simple trend with time, except for a suggestion of an increase (the percentage of incidents causing fatalities almost doubles) since about 1955.

The 533 fatal incidents recorded over the period 1600 to 2010 have caused a total of 278,880 fatalities. A small number of fatal incidents (Unzen in 1792, Tambora in 1815, Krakatau in 1883, Pelée in 1902, and Nevado del Ruiz in 1985) have caused 58% (162,928) of all fatalities, while just 13 incidents account for 74% (Figure 6). The largest five disasters are the only fatal incidents in the database to account for over 10,000 fatalities each, and the next five largest (Tungurahua in 1640, Grímsvötn in 1783, Kilauea in 1790, Santa María in 1902, and Kelut in 1919) are the only incidents responsible for between 5,000 and 10,000 fatalities (Figure 6). Single high-fatality incidents are most prevalent in Indonesia, followed by Guatemala. In spite of population growth increasing the number of people at risk, only three of the 13 highest-fatality incidents occurred after 1902, the date often cited as the birth of modern volcanology following the devastating eruption of Mont Pelée in Martinique (Lacroix, 1904).

The accumulative rate of increase in the number of fatalities shows large jumps as a result of the aforementioned handful of fatal incidents with large numbers of fatalities (Figure 7). We have removed these largest disasters from the data to investigate an underlying rate of accumulative increase that is less prone to their distorting effects. With the largest ten disasters removed, an average of 200 fatalities is recorded per year, though this increases with time from approximately 95 fatalities per year in 1600 to roughly 300 in 2010. Figure 8 shows the average number of fatalities per fatal incident over time; though noisy, the measure is generally decreasing over time. An average of over 300 fatalities per fatal incident was recorded in the 17th century, falling to below 100 in the 20th. These observations suggest that towards the start of the period, larger fatal incidents were better recorded than those causing few fatalities. Over time this recording bias has reduced, resulting in many more low-fatality incidents being recorded compared to high. Consequently, the average number of fatalities per year has risen with time, whilst the average number of fatalities per fatal incident has fallen.

The dominance of a small number of major disasters on historical fatalities is illustrated further in Figure 9. Whilst over 80% of fatal incidents are within the 0-10 and 11-100 fatality cohorts, fewer than 10% of fatalities are caused by such incidents; the opposite is true of the 1,001-10,000 fatality cohort. The largest five disasters are not included in Figure 9, though the overall distribution is similarly shaped but skewed towards the 10,001+ fatality cohort.

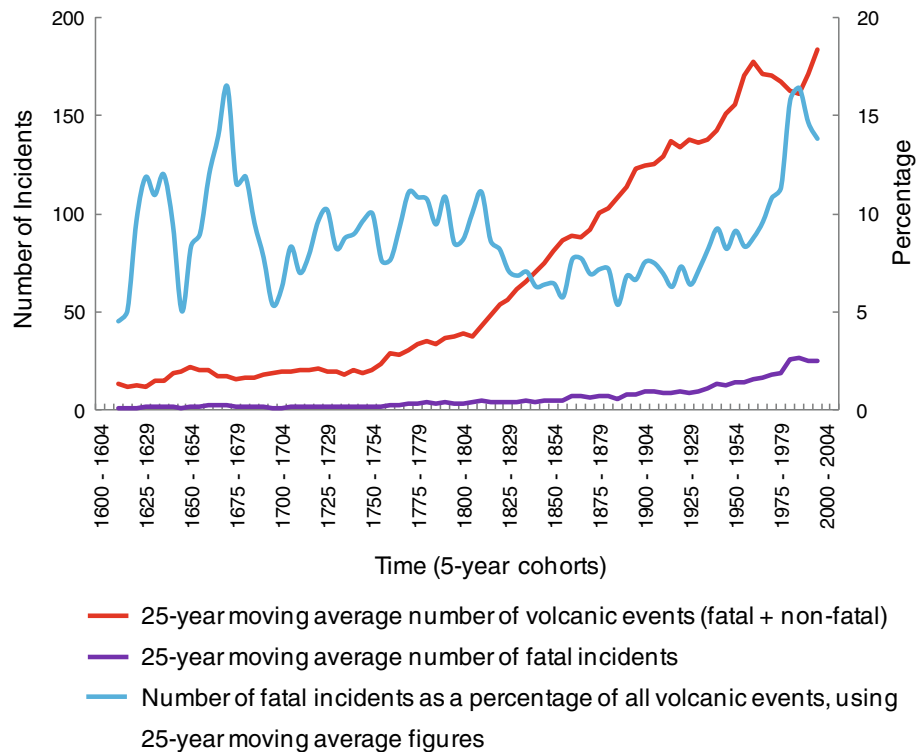


Figure 5 Time series of number of fatal incidents, volcanic events, and fatal incidents as a percentage of all volcanic events. Red line shows 25-year moving average of total number (both fatal and non-fatal) of volcanic events; purple line shows 25-year moving average of total number of fatal incidents; blue line shows 25-year moving average of the percentage of all events causing fatalities. Counts are calculated in five-year cohorts.

The fatalities distribution is evidently heavy-tailed (Figures 7, 8 and 9); a very large proportion of fatalities is attributed to just a handful of fatal incidents. The small sample size and statistically extreme nature of these fatal incidents necessitates their separate treatment and exclusion from some of our analyses. Doing so avoids distortion of our results, and allows for interpretation to be based on the statistically significant sample of frequently occurring fatal incidents (those with typically less than 1,000 fatalities). We make clear in the text and figure captions whether data have been excluded.

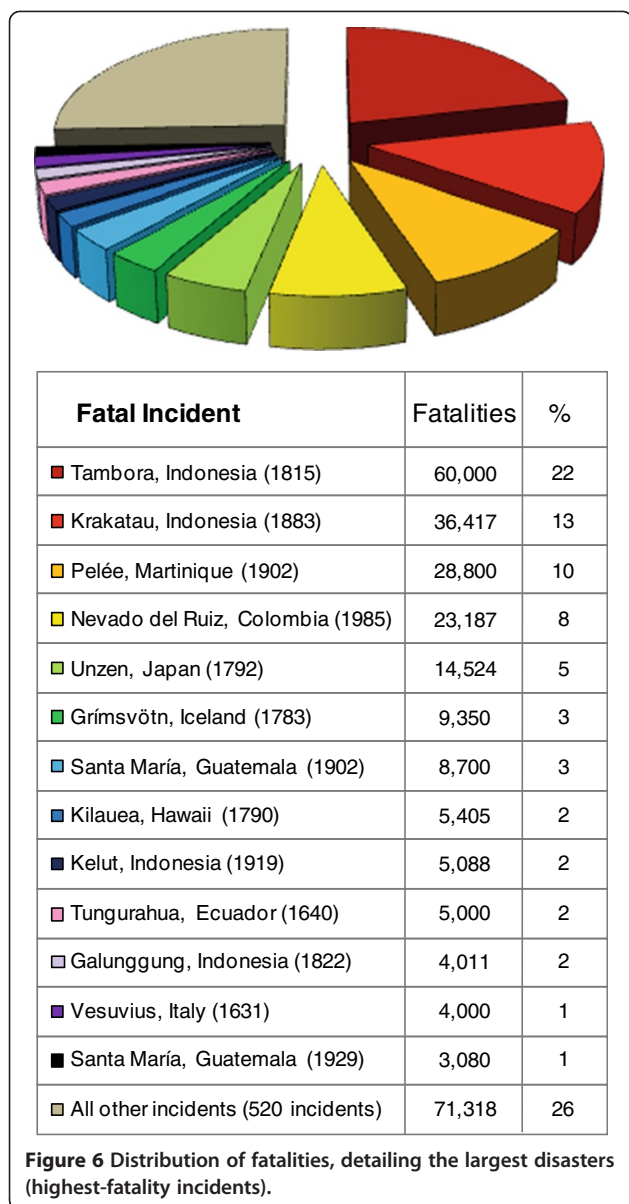
Analyses using fatality cause data

Following Simkin et al. (2001), we use 12 fatality causes to explore the spread of fatalities across different causal phenomena; these are defined in Table 1.

Data on fatality causes were only available in the SI and Witham databases. Coverage was incomplete and some data were quoted with uncertainty surrounding causes and the proportions of fatalities ascribed to them. The split of fatalities across the 12 causes was available for 441 of the 533 fatal incidents.

Using all the fatal incidents for which data were available (which includes the largest disasters), four causes – PDCs, indirect (predominantly starvation and disease), waves (tsunamis) and primary lahars – are identified as responsible for over 90% of the 274,501 fatalities (Figure 10a). When the largest five disasters are removed, these four causes dominate still, though to a lesser extent (responsible for 79% of fatalities). Tephra and secondary lahars have caused 13% of fatalities when the largest five disasters are removed (Figure 10b).

Fatality causes differ markedly across the spectrum of disaster sizes (Figure 11). Tephra and gas are common causes of fatalities in the zero to ten fatality cohort, with PDCs, tsunamis, lahars, and indirect fatalities the dominant causes for larger fatal incidents. The number of indirect fatalities is likely systematically under-represented in fatality totals though, due to the difficulty of quantifying and thus documenting such fatalities. The number of indirect fatalities is also likely decreasing with time as a result of the commencement of post-disaster relief through bi-lateral or multi-lateral efforts, which generally reduces the loss of life owing to starvation and disease. However, the data are insufficient for testing this notion.



Analyses using Volcanic Explosivity Index (VEI) data

With perfect recording, the magnitude-frequency relationship across the full spectrum of volcanic eruptions (both fatal and non-fatal) approximately follows a power law with exponent close to 1 (Deligne et al., 2010). However, volcanic eruptions and their magnitudes are not perfectly recorded over time; the probability of an eruption of given magnitude being recorded decreases back in time (Simkin, 1993; Coles and Sparks, 2006; Furlan, 2010), and the probability of an eruption occurring at a given time being recorded increases with magnitude (Deligne et al., 2010). The relationship between the frequency and magnitude of the fatality-causing subset of volcanic eruptions is further complicated by the likely correlation between fatalities and eruption magnitude. The relationship between

numbers of fatal incidents, numbers of fatalities, and eruption magnitude is here examined by dividing the data into Volcanic Explosivity Index (VEI) bins, using VEI (Newhall and Self, 1982) as a measure of magnitude. VEI data were added to the fatalities database from the SI Global Volcanism Program Volcano Reference Files.

With the largest five disasters excluded, the number of fatal incidents follows a log-normal distribution across VEI, with a mode of VEI 3. The number of fatalities is also log-normally distributed, but skewed towards larger magnitudes with a mode of VEI 4. Non-eruptive incidents and those with unknown VEI account for over 15% of all fatal incidents, and in combination they have caused 6,456 of the 278,880 total recorded fatalities (Figure 12a). We further explore this relationship by taking the ratio of the number of fatalities to the number of fatal incidents in each VEI cohort, with the largest five disasters again removed (Figure 12b); this procedure largely removes the effects of under-recording and any potential bias introduced by the largest disasters. The average number of fatalities per fatal incident is increasing with VEI, from VEI 0 incidents which cause an average of 49 fatalities, to VEI 6 incidents which cause an average of 1,906. The likelihood of an event causing fatalities is positively correlated with VEI (Figure 12c). Whilst the probability of a VEI 0, VEI 1, or VEI 2 eruption causing fatalities is well below 10%, eruptions of VEI 4 or greater have caused fatalities more often than not. All recorded VEI 6 eruptions have caused fatalities.

We consider the eruptions of Unzen (1792), Tambora (1815), Krakatau (1883), Pelée (1902), and Nevado del Ruiz (1985) separately. The two largest (in terms of fatalities) of these incidents, the eruption of Tambora (60,000 fatalities) and that of Krakatau (36,417 fatalities), were also the highest VEI (VEI 7 and VEI 6 respectively), and thus fit the overall relationship shown by the other 528 incidents. Figure 12b indicates that relatively few fatalities are caused by eruptions of VEI 3 and below. However, the incidents at Unzen and Nevado del Ruiz were both VEI 3 or below; the 1902 VEI 4 eruption of Pelée also caused more fatalities than would be predicted by VEI alone.

Taken as a whole, the data show that VEI is, in general, a good indicator of number of fatalities. However, closer examination of the fatalities record shows that there are some high VEI eruptions producing very destructive hazards that have caused very small numbers of fatalities, and vice versa; for these incidents, eruption magnitude (VEI) is not the dominant control on fatality counts, as previously noted by Blong (1988). Examples include the 1792 eruption of Unzen (VEI 2, 14,524 fatalities) and the 1912 eruption of Novarupta, Alaska (VEI 6, 2 fatalities). These two examples are easily explained; in the case of Unzen, the cause of fatalities was a tsunami related to the

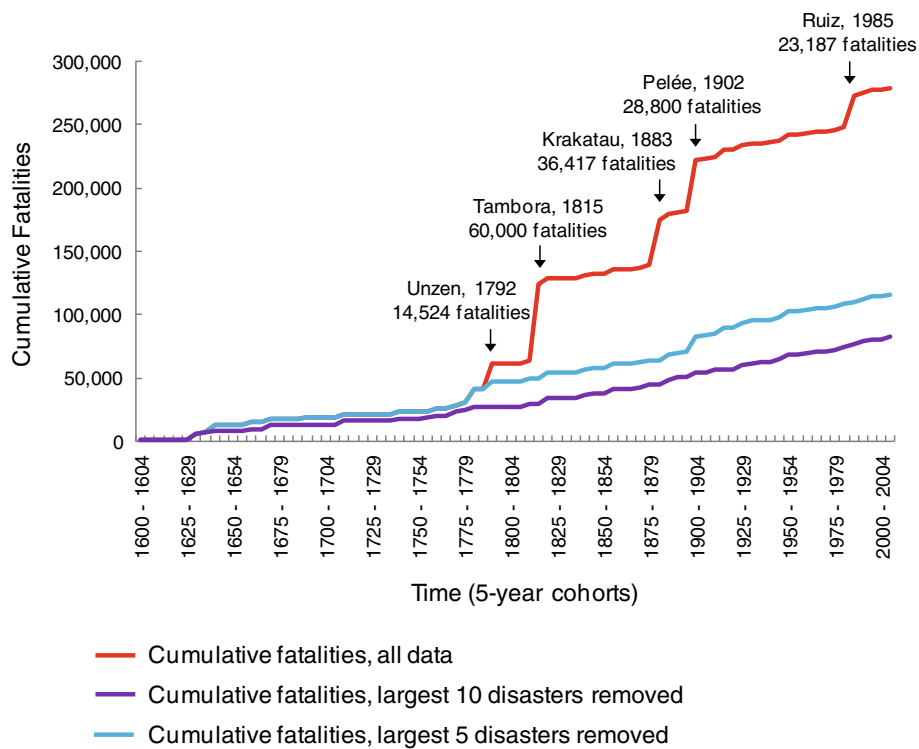


Figure 7 Cumulative number of volcanic fatalities. Shown using all 533 fatal volcanic incidents (red line), with the largest five disasters removed (blue line), and with the largest ten disasters removed (purple line). The largest five disasters are: Unzen, Japan, in 1792 (14,524 fatalities); Tambora, Indonesia, in 1815 (60,000 fatalities); Krakatau, Indonesia, in 1883 (36,417 fatalities); Pelée, Martinique, in 1902 (28,800 fatalities); Nevado del Ruiz, Colombia, in 1985 (23,187 fatalities). The sixth to tenth largest disasters are: Tungurahua, Ecuador, in 1640 (5,000 fatalities); Grímsvötn, Iceland, in 1783 (9,350 fatalities); Kilauea, Hawaii, in 1790 (5,405 fatalities); Santa María, Guatemala, in 1902 (8,700 fatalities); Kelut, Indonesia, in 1919 (5,099 fatalities). Counts are calculated in five-year cohorts.

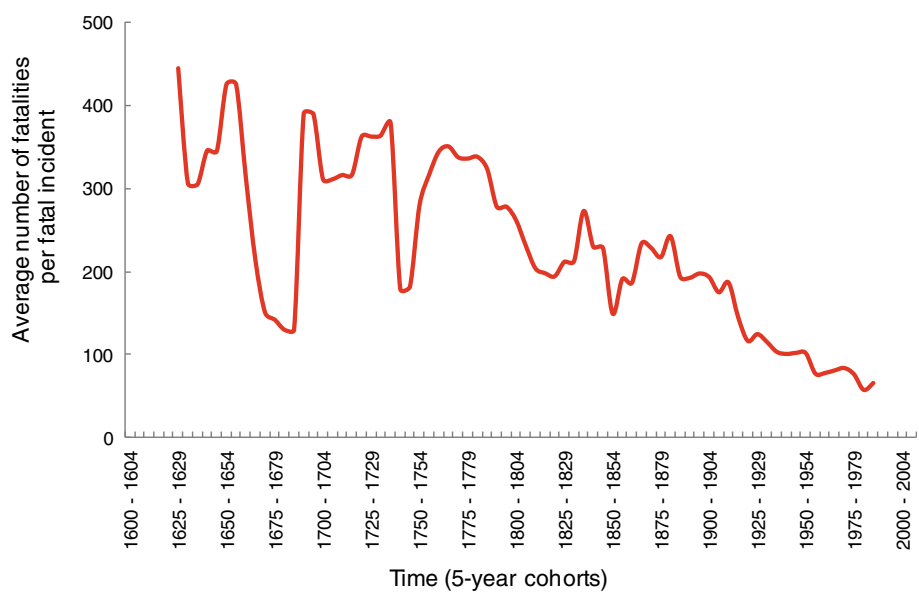


Figure 8 50-year moving average of the average number of fatalities per fatal incident. The largest ten disasters have been removed. Counts are calculated in five-year cohorts.

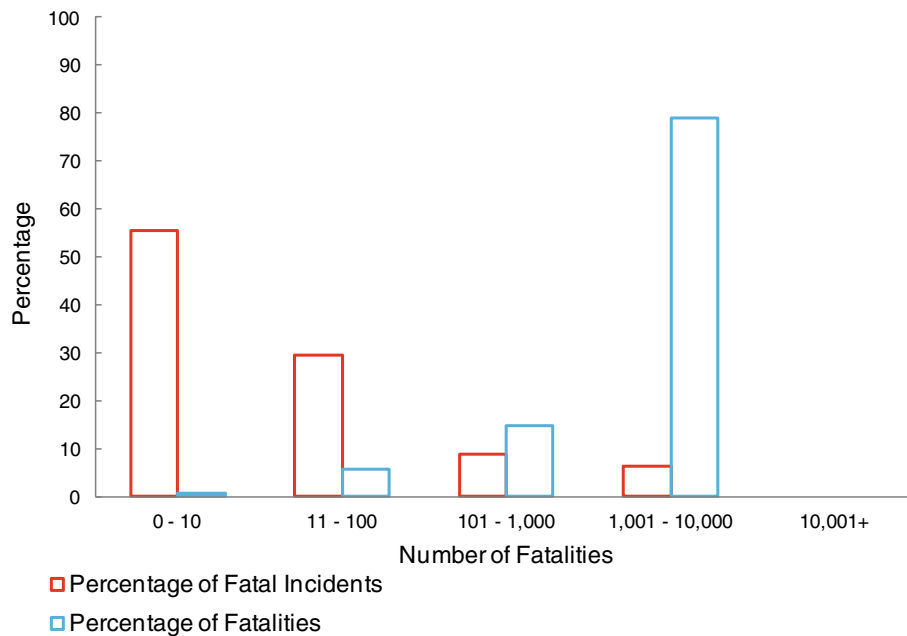


Figure 9 Distribution of fatalities and fatal incidents across fatality cohorts. The largest five disasters have been removed.

collapse of an old dome and thus unrelated to explosive magnitude, and at Novarupta there were very low numbers of people living in the region. We note that some high fatality eruptions are related to hazards such as lahars and tsunamis, which are not well correlated with VEI; this is discussed further below.

Analyses related to the spatial distribution of fatalities

Results relating to eruption size and style can only partially explain the variation in fatality figures seen across the database; we thus examine aspects of human exposure and vulnerability as alternative controls on numbers of fatalities. Some regions of the world appear more prone to fatalities than others (Figure 6). We explore these regional differences using the volcanic regional divisions used by IAVCEI (1951–1975) and adopted by the SI (Table 2). We have modified some of the region names in the figures and text for brevity and to focus on countries with volcanic fatalities.

Indonesia, South America, West Indies, Japan, and Mexico and Central America dominate the regional distribution of fatalities when all fatal incidents are included, as eight of the largest ten disasters are located within these regions. Fatalities in Indonesia account for almost half of all deaths from volcanic incidents, with four times more fatalities than the region with the second highest fatality count, South America (Figure 13). In Figure 14, where the largest ten disasters are removed, 38% of all fatalities still occur in Indonesia.

Taken together, Figures 13 and 14 show that the fatalities record in Indonesia is made up of both a few major

fatal incidents causing tens of thousands of deaths, as well as many smaller fatal incidents, meaning the region dominates the global fatalities record irrespective of the data subset used. The fatalities records of South America, the West Indies, Japan, Iceland, and Hawaii are dominated more strongly by single large disasters; at least 65% of each of these regions' fatalities are accounted for by the eruptions of Nevado del Ruiz (1985), Pelée (1902), Unzen (1792), Grímsvötn (1783), and Kilauea (1790), respectively.

Figures 13 and 14 ignore the differing numbers of volcanoes and events in each region. For example, large numbers of fatalities might be expected in Indonesia because the country has numerous volcanoes that erupt frequently and explosively, and a high population density. In order to factor out the number of events, we have normalised each region's fatality count by the total number of events (both fatal and non-fatal) occurring there, with the largest ten disasters removed, as shown in Figure 15. Higher ratios, to the left of the graph, denote regions more vulnerable to volcanic hazards and vice versa.

After accounting for differing numbers of events across the regions, the populations of the West Indies, Philippines and SE China, Mediterranean, Indonesia, and Melanesia emerge as the most vulnerable to volcanic hazards; each has an average of 25 or more fatalities per volcanic event. North America (composed of Canada, Alaska, and the conterminous USA), Indian Ocean, and Kuril Islands and Kamchatka have the lowest numbers of both absolute fatalities (Figure 14) and fatalities relative to number of events (Figure 15). Each of these regions, along with

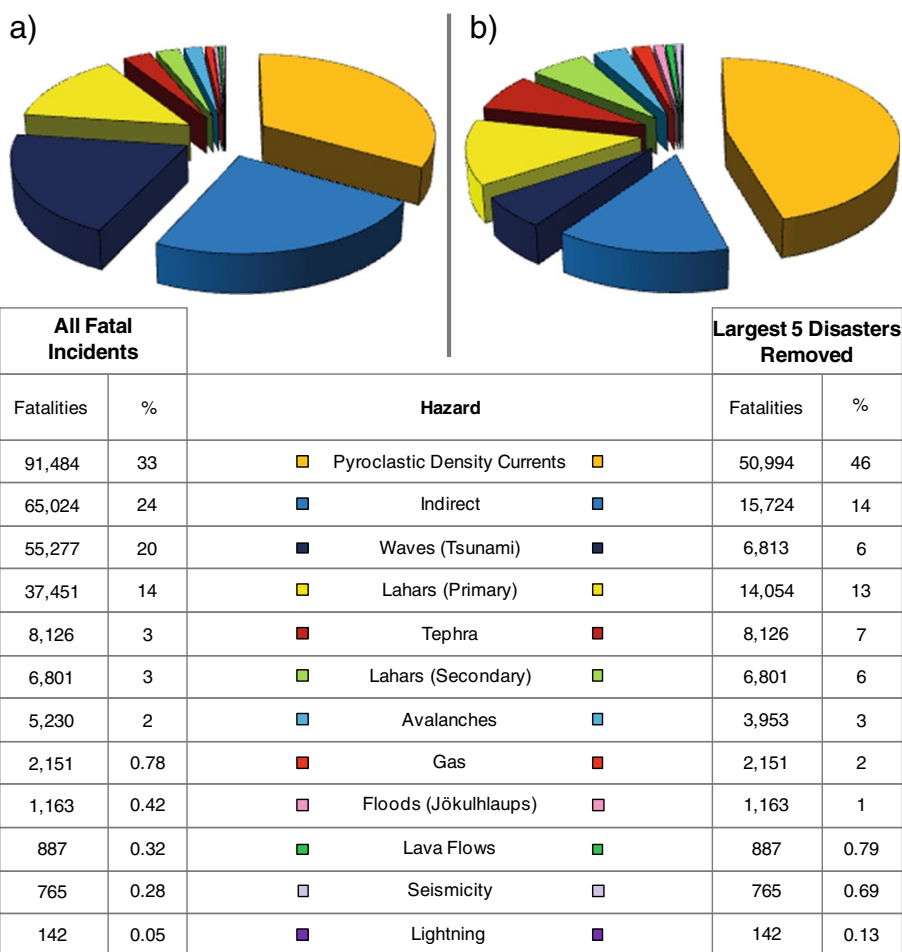


Figure 10 a) Distribution of fatalities across fatality causes, for all fatal incidents; b) Distribution of fatalities across fatality causes, with the largest five disasters removed.

Iceland, has an average number of fatalities per event (non-fatal and fatal) below 0.5.

The average number of fatalities per volcanic event spans three orders of magnitude, from 0.12 (Kuril Islands and Kamchatka, and Iceland) to 201 for the West Indies. There are significant differences between regions in the probability of an event causing fatalities, and the likely number of fatalities.

The spatial distribution of fatalities can be examined at a finer scale by exploring how numbers of fatalities vary with distance from the volcano. A written incident description that includes such distance data was available for 31 fatal incidents across the SI and Witham databases. These 31 fatal incidents included the eruptions of Tambora and Nevado del Ruiz, which were removed from the sample. Incidents for which indirect causes (famine and disease) were responsible for the majority of fatalities were also removed, as it is not possible to give a well-defined estimate of hazard extent in these cases. A consequently small

sample of 27 incidents is thus plotted in Figures 16a and b, with data points in Figure 16a colour coded according to the dominant cause of fatalities.

Unsurprisingly, there is a trend of increasing numbers of fatalities with distance from the volcano. Most fatalities (over 85%) are concentrated between distances of 5 km to 30 km from the volcano (Figure 16b), likely due to low population densities in the immediate vicinity of volcanoes; within 5 km, terrain is often too steep or frequently impacted for habitation, or designated as national parks, and thus only visitors tend to be present within 5 km. Six of the seven fatal incidents that caused fatalities one kilometre or less from the volcano claimed fewer than 41 lives, whilst six of the ten that caused fatalities ten or more kilometres from the volcano killed over 140 people. Fatalities closest to the volcano are most commonly caused by tephra, and those more distant are commonly caused by PDCs, primary and secondary lahars, and waves (tsunamis) that have affected

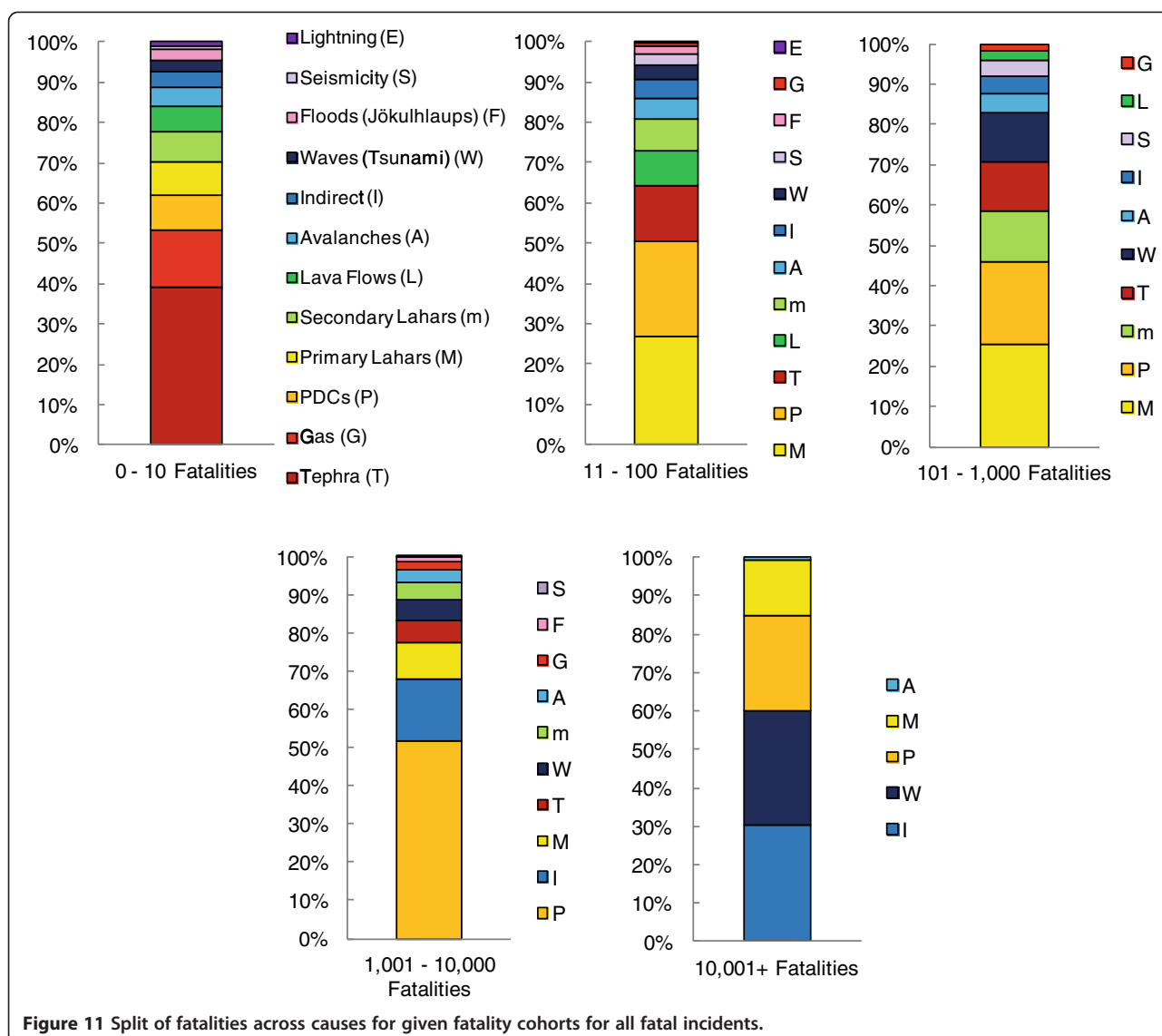


Figure 11 Split of fatalities across causes for given fatality cohorts for all fatal incidents.

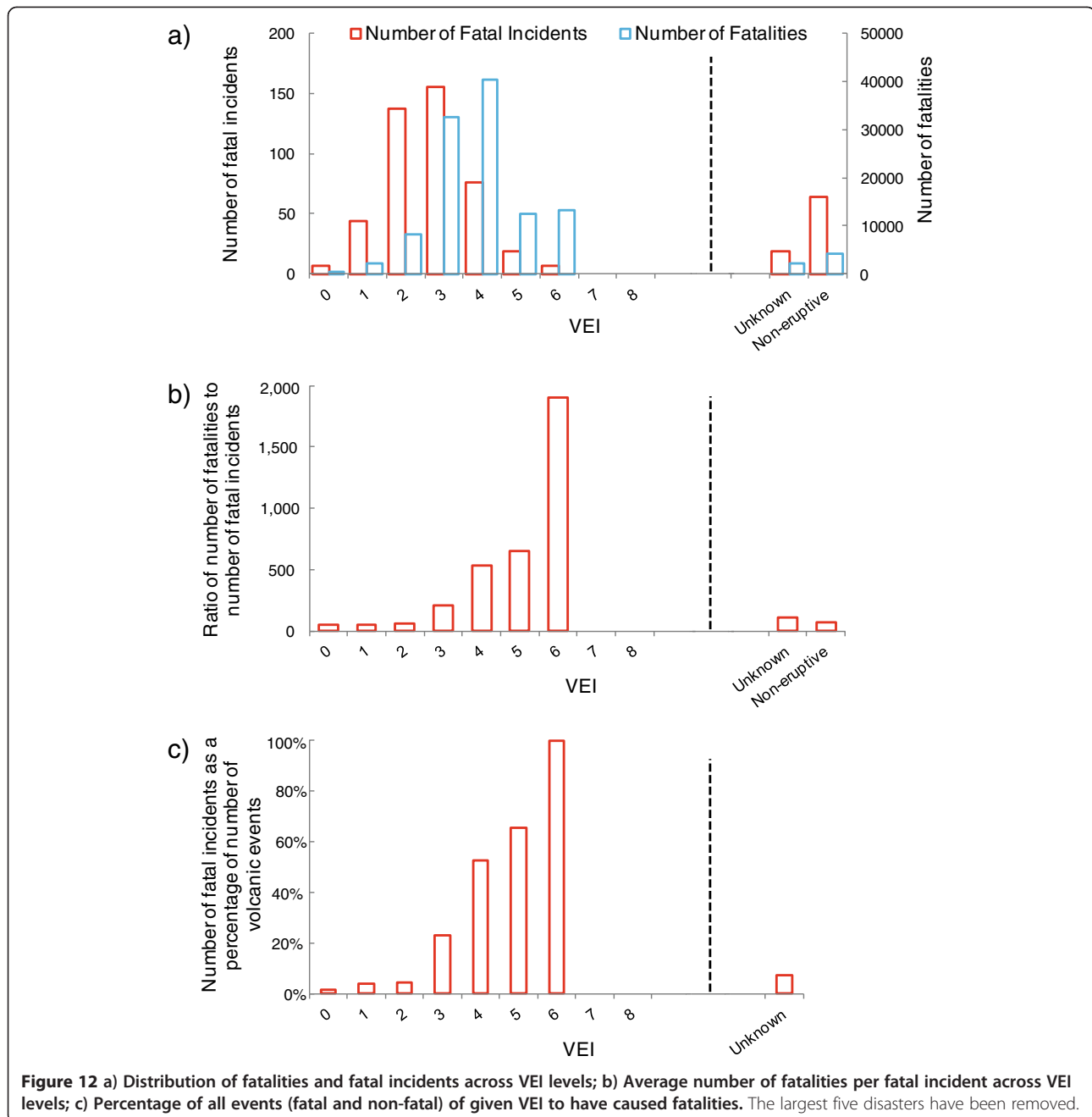
areas with permanent residents. Particularly for the latter of these fatality causes, distance data are likely approximate as those killed are rarely concentrated in one single location, and may in fact be spread over some kilometres. Note that the largest disasters caused fatalities far beyond 10 km from the volcano.

Analyses using Human Development Index (HDI) data

We note from Figure 15 an apparent link between a region's level of development and number of volcanic fatalities. Most of the regions which lie above the 38-country average number of fatalities per volcanic event are classed by the United Nations (UN) as developing, whilst most below are classed as developed (UN Statistics Division, 2011). We therefore investigate the relationship between regional development status as measured by the Human Development Index (HDI) and number of volcanic fatalities in

finer detail (Figure 17). Regional HDI values were calculated by averaging HDI figures over time and constituent countries with volcanic fatalities. Following the UN, we classify regions' development levels as low, medium, high, or very high, and these are coloured red, orange, yellow, and green, respectively, in Figure 17.

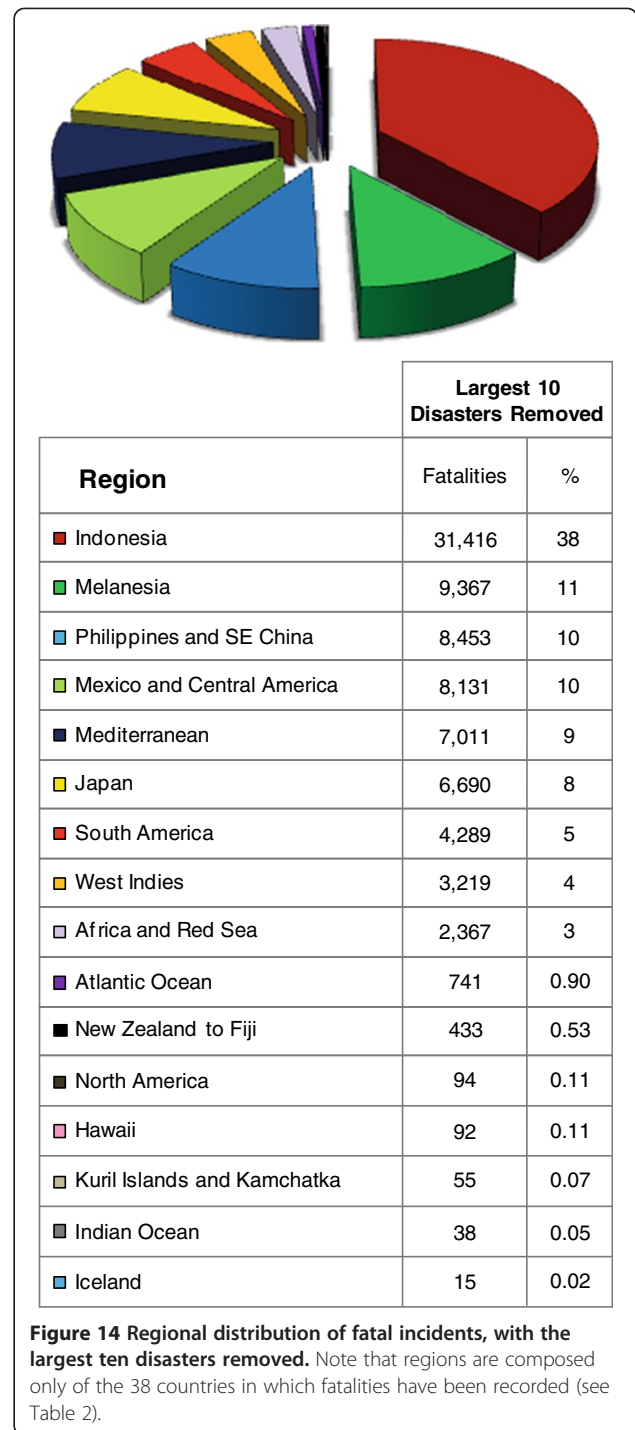
Unlike Witham (2005), we see no simple relationship between a region's HDI level and average number of fatalities per event (Figure 17); the correlation between regional HDI value and average number of fatalities per event is not statistically significant. However, the measure is on average much higher for low, medium, and high HDI regions than for those classed as very high. The weakness of this relationship may be a reflection of the fact that a small number of regions are composed of countries with varied HDI levels, such as New Zealand and Tonga, which form the New Zealand to Fiji region.



Analyses using population data

Lastly, we explore changes to human vulnerability, defined as the likelihood of a certain proportion of the population living in the footprint area of a volcano being killed, over time. Rapid population growth, particularly in densely-populated urban centres located close to volcanoes, has raised the number of people at risk from volcanoes (Dilley et al., 2005; UNFPA, 2011; Small and Naumann, 2001); however, developments in volcanology and proactive implementation of mitigation measures have attempted to counter this.

We investigate the balance of these factors' effects on society's vulnerability to volcanic hazards, as measured by the number of fatalities, by scaling by population. We are concerned only with the 38 countries in which volcanic fatalities have been recorded; this avoids any distortion that may be introduced through the inclusion of large populations that are not threatened by volcanic hazards (such as those of India and Brazil). In addition, the largest ten disasters are removed. We then define the Volcano Fatality Index (VFI) as the number of fatalities divided by the product of the number of recorded volcanic events



(fatal and non fatal) and population, in a fixed time period. This value is then multiplied by 100,000 to give a rational range, with a 20th century value of 0.8 (almost 1). These normalisations should largely remove the effects of under-recording and population growth on our results. If vulnerability to volcanic hazards, as measured by fatalities, was only controlled by population and numbers of events, then the VFI should be approximately independent of time if vulnerability remained unchanged. In Figure 18,

the VFI is calculated as a 25-year moving average (using five-year cohorts). Century-wide averages are also displayed to help elucidate trends.

Despite much noise in the data, we observe that the VFI declines markedly with time. Values are high and variable in the 17th and first half of the 18th centuries (average VFI of 19.7 and 9.4, respectively), then fall sharply over the latter half of the 18th century. An

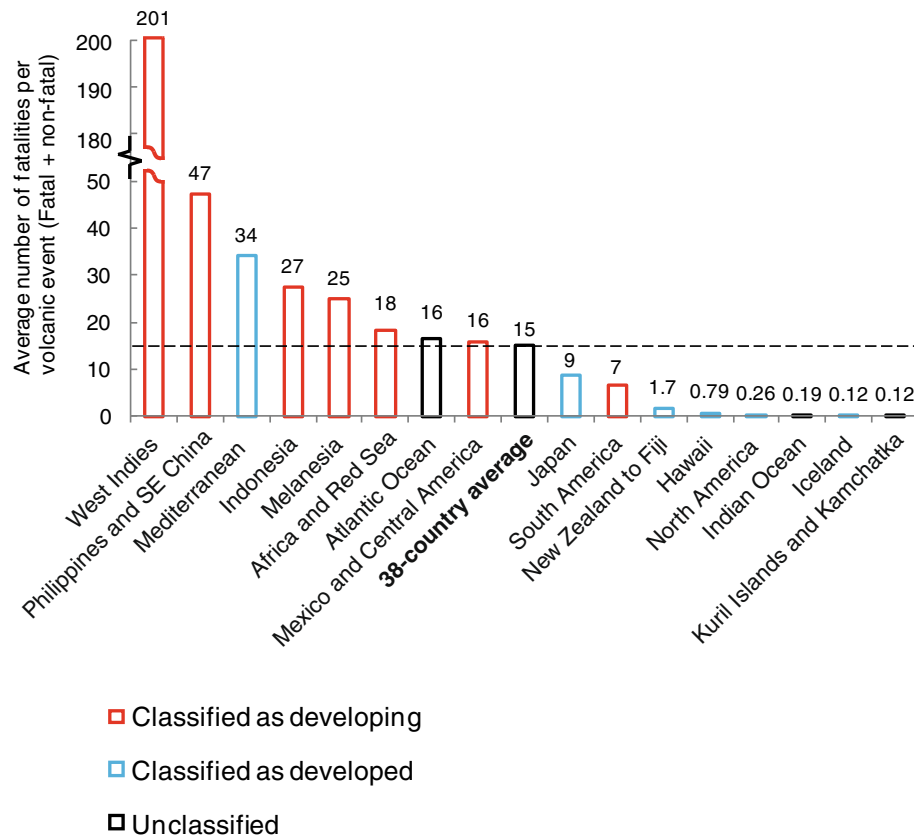


Figure 15 Average number of fatalities per volcanic event (fatal and non-fatal), organised by region. Colour coding is used to show development level, as classified by the UN. Black dashed line shows 38-country average number of fatalities per volcanic event. The largest ten disasters have been removed. Note that regions are composed only of the 38 countries in which fatalities have been recorded (see Table 2).

approximately steady decrease in VFI is observed from 1800 onwards. The average 19th century VFI is 2.8, reducing to an average VFI of 0.8 in the 20th century and 0.4 across the last three decades of this study, although it is acknowledged that this is too short a time period to be truly comparable.

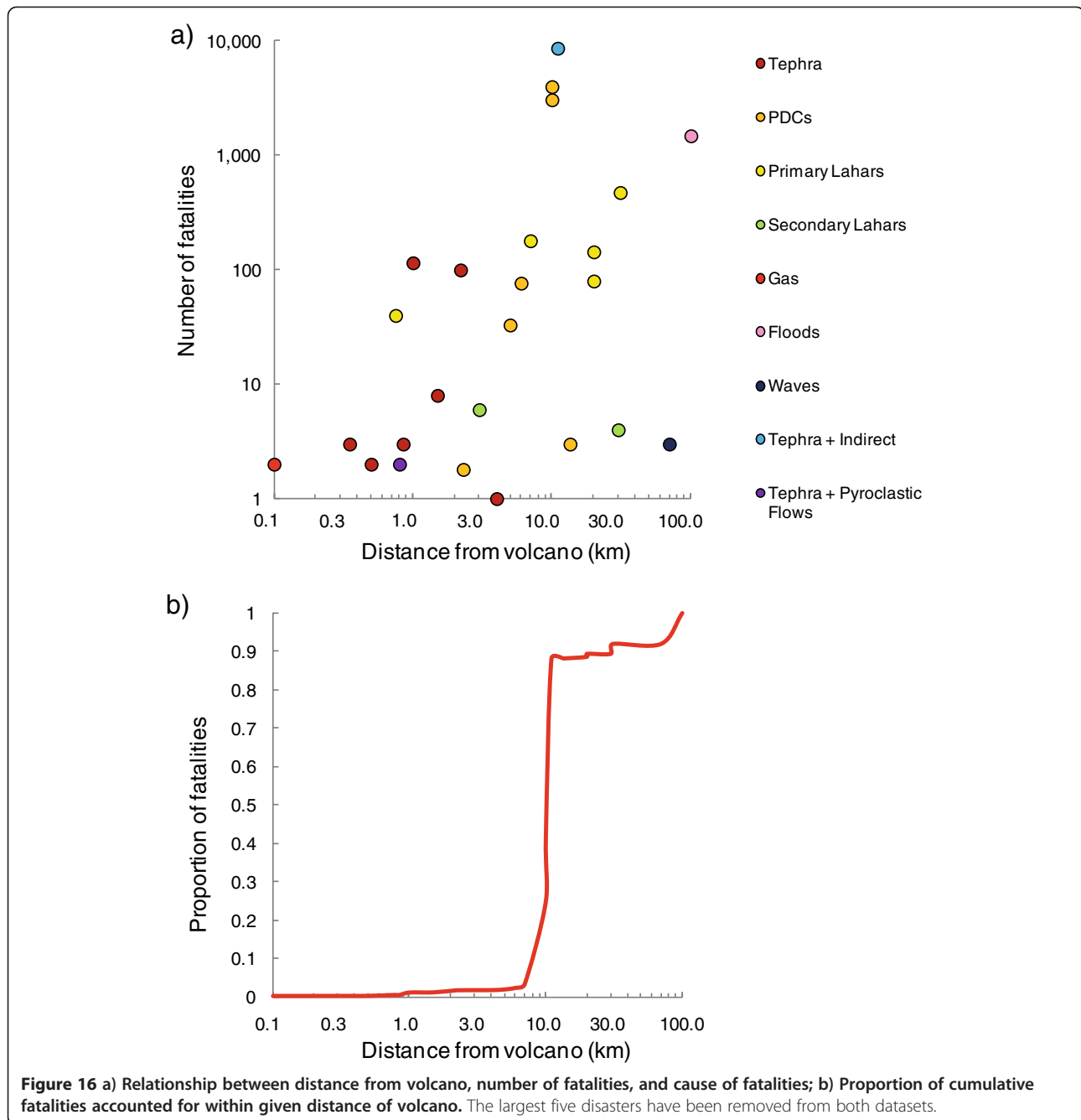
Discussion

The historical record of volcanic fatalities provides a major source of evidence on the causes of volcanic disasters. Fatality data are an indicator of the relative importance of different kinds of volcanic hazards and the role of eruption magnitude and style, and provide insights into social, economic, and physical vulnerability of populations. However, to use these data requires recognition of their many biases, uncertainties, and incompleteness, as well as other more minor inaccuracies (see Hittelman et al., 2001). The data are also influenced by many factors which change with time, such as the way in which disasters are reported, population growth, advances in modern science, and changes in societal resilience.

The database lists 533 fatal incidents spanning a 410-year period; an estimated 278,880 fatalities have been recorded,

caused by incidents at 198 different volcanoes spread across 38 countries. Our estimate of the total number of fatalities is well aligned with the numbers recorded in similar studies by Tanguy et al. (1998), Witham (2005), and Simkin et al. (2001). We have established that the non-systematic uncertainties are statistically within a factor of three.

One central difficulty in the interpretation of volcanic fatality data, and likely fatality data pertaining to other natural hazards, is the major influence of a small number of incidents causing thousands of fatalities. Although the total number of recorded fatal incidents is 533, just five incidents have caused 58% of reported deaths and just ten have caused 70%. Inferences based on total fatalities are likely to have significant uncertainties related to the small sample size at the extreme tail of the distribution. Furthermore, the sample of fatalities data is drawn from a time series that is likely non-stationary and strongly affected by time dependent biases. As such, we have also presented data in terms of numbers of fatal incidents and by leaving out the largest disasters, under the assumption that the many incidents with fewer fatalities are more representative of trends and distributions.



Irrespective of how the data are treated, lahars and PDCs emerge as the main cause of fatalities. However, the impacts of some hazards are very dependent on how the data are treated; for example, volcanogenic tsunamis become more prominent as a consequence of including the 1792 Unzen and 1883 Krakatau eruptions. Some of the hazards, such as tephra and lightning, do not result in mass casualties. Amongst the most frequent incidents though (those causing zero to ten fatalities, which account for 55% of all fatal incidents) tephra and gas are

the most common causes of fatalities. Secondary effects such as famine and disease account for significant numbers of fatalities (24%), but the reliability of these data is likely less than that of data pertaining to primary hazards.

We use VEI, the most easily accessible measure of eruption magnitude, to explore the relationship between fatalities and eruption size. The data show a tendency for fatality numbers to increase with VEI, and the probability of fatalities increases markedly with VEI. There are,

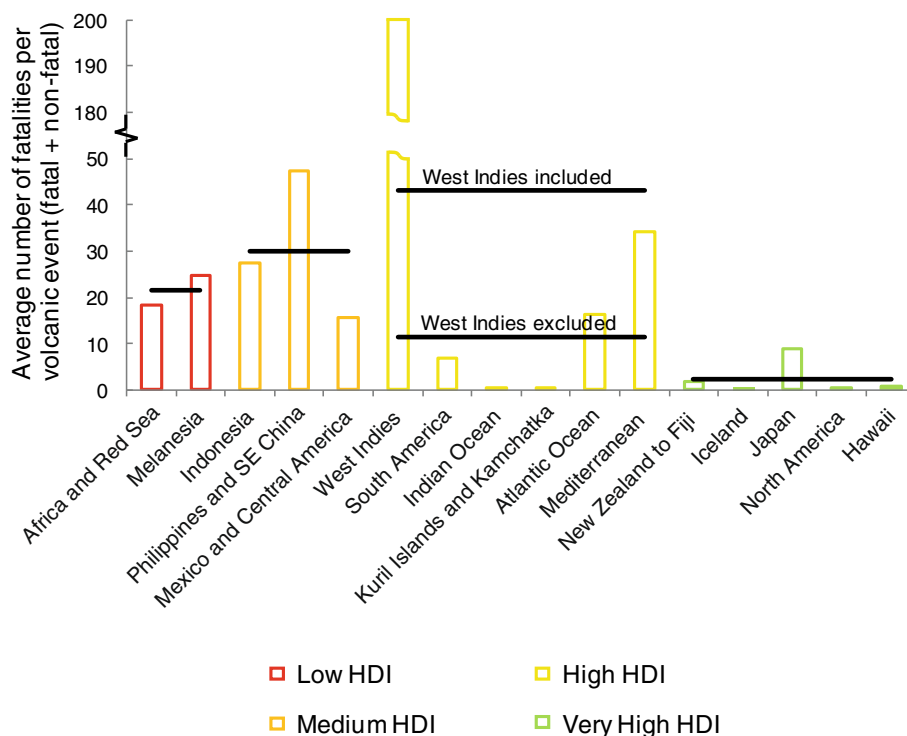


Figure 17 Average number of fatalities per event (fatal and non-fatal) organised by region. Regions are listed in increasing HDI level from left to right; colour coding is used to show HDI level, and black lines display HDI-level averaged data. The largest ten disasters have been removed. Note that regions are composed only of the 38 countries in which fatalities have been recorded (see Table 2).

however, some high fatality disasters for low VEI eruptions. One reason for the imperfect relationship with VEI is that some major hazards, namely PDCs and lahars, are not always strongly related to VEI. This is exemplified by eruptions such as those of Nevado del Ruiz in 1985 (VEI 3) and Pelée in 1902 (VEI 4). The severity of lahars is most closely linked to the volume of water at or near the site of eruption, in the form of a crater lake, ice cap, or precipitation; large volumes of water can be rapidly created from melted ice caps following fairly low magnitude eruptions. Hazard footprints for PDCs are not simply related to VEI, an index derived from estimates of eruptive volume, as their hazard footprints are determined by other factors such as eruption intensity and flank topography, as well as volume.

Population exposure to hazard – here defined as the number of people living within the potential hazards footprint of a volcano – and population distribution are other major controls on volcanic fatalities. Our data show that the number of fatalities is related to hazard extent (Figures 16a and b), and highly destructive hazards with typically large hazard footprints, such as PDCs, lahars, and tsunamis, have caused the greatest numbers of fatalities. A large proportion of fatalities have occurred at distances between 5 km and 30 km from the

volcano, which has grave implications for various cities around the world, as depicted in Figure 1. Further, these results lend support to methods using populations within 10 km and 30 km radius circles as measures of exposure (e.g. Ewert and Harpel, 2004; Siebert et al., 2010). The numbers of fatalities at different extents can also be used to enable more detailed assessments of population exposure to volcanoes, such as the Population Exposure Index of Aspinall et al. (2011). A large fraction of the world's fatalities have occurred in the densely populated countries of Maritime Southeast Asia (Indonesia and the Philippines) and Mexico and Central America (Mexico, Guatemala, El Salvador, Nicaragua, Costa Rica), which are predominantly located in the tropics. These are countries identified by Ewert and Harpel (2004) and Small and Naumann (2001) as having a high Volcano Population Index (VPI; numbers of people living near to volcanoes), and higher population densities close to volcanoes than most other volcanically active regions. Unfortunately, from a hazards perspective, many of these regions are also those with high levels of volcanic activity (Siebert et al., 2010).

The data suggest some broad negative correlation between a region's HDI and number of fatalities. Some of the weakness of this relationship may be in part attributable to the difficulty of calculating regionally representative HDI

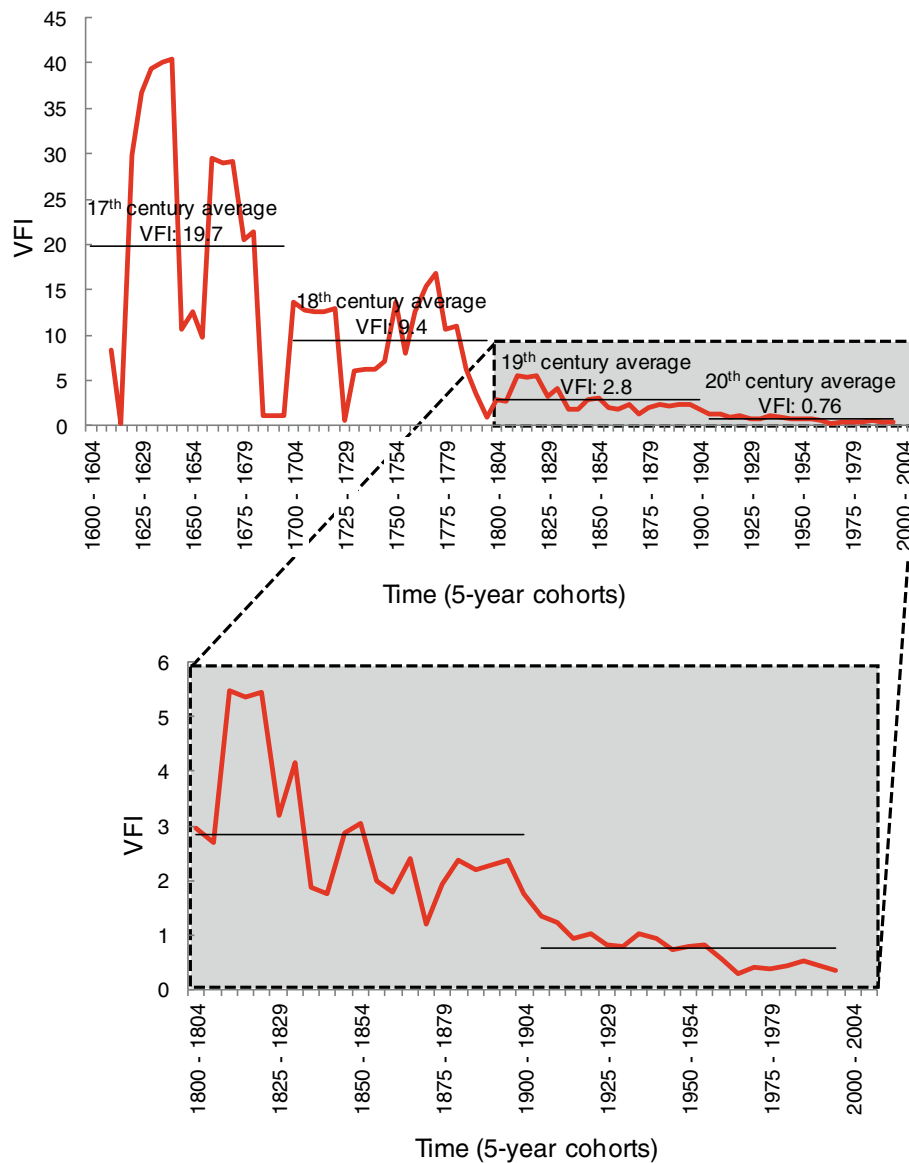


Figure 18 25-year moving average Volcano Fatality Index (VFI). Century-wide average VFI figures are displayed in black. Inset plot shows 25-year moving average VFI over the period 1800 – 2010. The largest ten disasters have been removed.

figures for some regions (e.g. New Zealand to Fiji, composed of New Zealand and Tonga); also, the relatively low average numbers of fatalities per event in very high HDI regions may be a reflection of better recording of events with few fatalities than in less developed regions. In spite of this, the data show that some of the regions with highest numbers of fatalities are composed of low or medium HDI island nations or archipelagos (e.g. Indonesia, the West Indies). Developing island populations seem particularly vulnerable to volcanic hazards, likely because of their inherent proximity to a volcano combined with the increased difficulty of evacuating by sea or air. Volcanic tsunamis are also confined to volcanic islands or volcanoes near coastal areas.

Climate is another factor; many fatal incidents have occurred in tropical latitudes, likely for two reasons. Firstly, and most importantly, cooler temperatures at high altitudes on the flanks of volcanoes have led to large populations settling very close to volcanoes compared to areas of more temperate climate. High VFI figures in tropical volcanic areas support this assertion (Ewert and Harpel, 2004; Siebert et al., 2010). Secondly, high rainfall in the tropics makes highly-destructive lahar hazards more probable. The data show that 58% of fatal incidents, but 85% of fatalities, have occurred within the tropics, and over 70% of all incidents for which lahars are listed as a cause of fatalities have occurred there.

We have made some necessary simplifications and applied normalisation procedures designed to correct for data biases and distortions. There are major problems with under-recording of historical datasets on volcanism (Simkin, 1993; Coles and Sparks, 2006; Furlan, 2010; Siebert et al., 2010). Since 1600 AD, the number of recorded events has increased significantly, although it was not until the late 19th century that recorded eruptions reached a rate of 50% of current eruption rates (Siebert et al., 2010). The number of fatal incidents as a percentage of all events fluctuates markedly between 6% and 14%, with no systematic trend. However, the number of fatalities per fatal incident declines from an average of over 400 to less than 100 between 1600 and 2010. We attribute this trend to a change in reporting; in the 17th century volcanic events were likely recorded only if they led to a human disaster with mass casualties, but systematic, scientific recording was limited. We suggest that the high numbers of fatalities per fatal incident toward the beginning of the study period is an artefact of the many missing incidents with small numbers of fatalities. The decreases in the average number of fatalities per fatal incident and VFI over time (Figures 8 and 17) are most significant from the late 1700s onwards. Such changes suggest that the advent of rational scientific thought has initiated recording of eruptions as natural events rather than just disasters; modern science and the idea that one might report volcanic eruptions and their effects as a systematic endeavour emerged in earnest during the 19th century (Siebert et al., 2010). Furlan (2010) estimates that the percentage of all volcanic events recorded prior to 1900 is 40%. We thus offer a crude estimate of the effects of under-recording on the number of fatalities by assuming that in the first 300 years of the study, under-reporting of all events was proportional to the under-reporting of fatalities, with the exception of the very largest disasters which are likely fully reported. In this case, the number of fatalities between 1600 and 1900 would increase by more than 100,000.

Population growth in the 410 years of our study has been exponential. The proposed VFI is a way of correcting for this population growth as well as under-recording. If there had been no change in vulnerability – the likelihood of a certain proportion of the exposed population being killed – with time then the VFI should be approximately constant. We observe high and variable VFI values across the 17th and first half of the 18th centuries, likely due predominantly to poor data quality (related both to volcanism and to population); the reduction in the VFI over this period is probably therefore attributable to changes in both the level of reporting and its accuracy. Some of the variation may also be due to natural variation in eruption frequency, but this is not distinguishable within the time-frame of this study.

However, the VFI has reduced dramatically, by roughly one order of magnitude, in the 20th century, suggesting a real reduction in society's vulnerability to volcanic hazards. Key explanatory factors include improvements in scientific hazards assessments and early warning systems, timely evacuations, better preparedness, and greater population awareness. The first volcano observatories were founded in the second half of the 19th century and early 20th century. The "birth" of volcanology in 1902 from a hazards mitigation perspective followed the pivotal eruptions of Pelée and Soufrière St Vincent in the eastern Caribbean. Planned mass evacuations as a means to mitigate disaster potential started in earnest in the 1950s and 1960s (Witham, 2005). More recently there have been extensive technological and scientific developments, which include the inaugural use of remote sensing in volcano monitoring in the 1980s following the 1980 eruption of Mount St. Helens (Tilling and Lipman, 1993; Newhall, 2000; Ewert and Swanson, 1992; Sparks et al., 2012). This is *prima facie* evidence that science has had a major beneficial impact on volcanic disaster risk reduction.

One of the major difficulties in disaster risk reduction is demonstrating conclusively that investments in mitigation, civil protection, and preparedness, informed by science through early warning and hazards assessments, have saved lives. Our evidence is largely qualitative and anecdotal, though the proposed VFI enables a more quantitative approach. First we observe that VFI has decreased from about 1.5 to 0.3 in the last 100 years, since the inception of modern volcanology from a hazards perspective. A simple postulate then is that the VFI would have remained at about 1.5 throughout the 20th century had no progress been made in disaster and risk reduction strategies. This value can be used to estimate the fatalities that would have potentially happened as a consequence of increased exposure of society due to exponential population growth. We estimate this number as 85,269, compared to the actual number of 31,728. These figures are calculated with the eruptions of Pelée and Nevado del Ruiz excluded, which caused a further 51,987 fatalities in combination.

While these evaluations are very approximate and uncertain, they do provide evidence that the science of volcanology has been successful in reducing volcanic disaster risk. Several recent eruptions exemplify these successes with timely evacuations of thousands to tens of thousands of people from high hazard zones shortly before major eruptions. In 1991, about 20,000 people were evacuated from the flanks of Pinatubo, Philippines, prior to the cataclysmic VEI 6 explosive eruption on 15th June that produced massive PDCs that devastated these evacuated areas (Newhall and Punongbayan, 1996). If the populations surrounding Pinatubo had not been evacuated, the eruption may well have formed one of the largest disasters

in our database. Several thousand people were evacuated from the flanks of Soufrière Hills, Montserrat, in 1996 prior to many PDCs and a major volcanic blast (VEI 3). Most recently, over 10,000 people were evacuated from areas surrounding Merapi, Indonesia, which were later destroyed by PDCs during the VEI 4 October to November 2010 eruption (Suroño et al., 2012). Other successful evacuations include those at Banda Api (1988), Rabaul (1994), Tungurahua (2006), and Nevado del Huila (2007, 2008) which, in combination, saved tens of thousands of lives. The approximately 50,000 lives saved according to calculations based on VFI is therefore likely an underestimate.

Conclusions

With hundreds of millions of people potentially at risk from volcanic hazards, developing a bank of data covering volcanic events and their impacts is a very useful resource in assessing and managing volcanic risk in the future. Our database enhances this collection of information, and contains records of more fatal incidents than other works before it.

Our database spans the years 1600 AD to 2010 and details 533 incidents that have caused a combined total of 278,880 fatalities, at 198 volcanoes across 38 countries. Whilst most incidents cause ten or fewer fatalities, the largest five disasters – those of Unzen, Tambora, Krakatau, Pelée, and Nevado del Ruiz – have each caused over 10,000. The regions in which these incidents occurred dominate the regional split of fatalities; when these disasters are excluded from analyses, Indonesia, Melanesia, the Philippines and SE China, the West Indies, and Mexico and Central America have seen the greatest human loss from volcanism. These findings are attributable to various factors, such as tropical latitude, high population density close to volcanoes, and location of volcanoes on low and medium HDI islands.

PDCs, followed by indirect causes, tsunamis, and primary lahars, have caused most fatalities, though tephra and gas are the most common causes in low-fatality (ten or below) incidents. The average number of fatalities per fatal incident and the probability of fatalities are broadly correlated with VEI. There are likely weaknesses in this relationship because the extent and intensity, and therefore number of fatalities, resultant from lahars and PDCs are not strongly dependent on VEI, and such hazards have caused a combined 50% of all fatalities. There is some evidence that a region's development level is also a possible explanator of the number of fatalities, with very high HDI regions experiencing fewer fatalities per event on average.

The VFI, a measure of fatalities that aims to remove the effects of both population growth and under-reporting, provides a new approach to examining changes in vulnerability to volcanic hazards over time. Whilst the commencement of recording of volcanic eruptions as natural

events rather than solely as disasters may explain the steady decrease in the VFI seen in the 17th and 18th centuries, the five-fold decrease seen from approximately 1800 onwards is likely a result of real decreases in populations' vulnerability to volcanic hazards. Such decreases are generally temporally linked to developments in the science of volcanology and improvements in societal resilience, and thus suggest the benefits of investments in disaster risk reduction are borne out by the fatalities record. Using the VFI, we estimate that improvements in hazard management as a whole have saved approximately 50,000 lives during the last 100 years. This figure is likely an underestimate, as the numbers of lives saved through evacuations prior to eruptions in the 1980s, 1990s, and 2000s far exceeds 50,000.

Several issues with the underlying fatalities data have been identified: potential bias introduced by analysis of the fatalities population as a whole due to its heavy-tailed nature; under-recording of fatalities and fatal incidents, as well as of volcanic events as a whole; uncertainty in fatality figures. We have attempted to overcome the former two issues by excluding the largest disasters and normalising by numbers of fatal incidents, respectively, but note that under-recording still affects volcanic event and fatality records. The scope of analyses undertaken and interpretations made using the fatalities data are limited by these issues.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MA acquired data, participated in the study's design, coordinated the project, carried out all statistical data analyses, and drafted the manuscript. RSJS conceived of the study and helped draft the manuscript. LS and HSC acquired data and participated in the study's design. JE participated in the study's design. All authors have read, reviewed, and approved the manuscript.

Acknowledgements

This research was supported by the European Research Council VOLDIES grant (MA, RSJS, HSC) and the US government (LS, JE). We thank Susanna Jenkins for careful review of an early version of the manuscript, and Russell Blong and an anonymous reviewer for their comments.

Author details

¹Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK. ²Global Volcanism Program, Smithsonian Institution (Ret.), Washington, DC, USA. ³USGS, Cascades Volcano Observatory, Vancouver, WA, USA.

Received: 25 July 2012 Accepted: 8 February 2013

Published: 14 February 2013

References

- Aspinall W, Auker M, Hincks TK, Mahony S, Nadim F, Pooley J, Sparks RSJ, Syre E (2011) Volcano hazard and exposure in GDRFF priority countries and risk mitigation measures - GFDRR Volcano Risk Study. Bristol University Cabot Institute and NGI Norway for the World Bank: NGI Report 20100806, Bristol, 3 May 2011
- Blong RJ (1984) Volcanic hazards. Academic Press, Sydney
- Blong RJ (1988) Assessment of eruption consequences. Proceedings Kagoshima International Conference on Volcanoes, Kagoshima
- Blong RJ (2003) A review of damage intensity scales. *Nat Hazard* 29:57–76

- Bulletin of Volcanic Eruptions, Nos. 1 – 33 (1960–1996) Volcanological Society of Japan, Tokyo
- Centre for Research on the Epidemiology of Disasters (2010) EM-DAT: The OFDA/CRED International Disaster Database. Université Catholique de Louvain, Brussels, Belgium, Available online: www.emdat.net. ISBN EM-DAT: The OFDA/CRED International Disaster Database
- Chester DK, Degg M, Duncan AM, Guest JE (2001) The increasing exposure of cities to the effects of volcanic eruptions: a global survey. *Global Environmental Change Part B: Environmental Hazards* 2:89–103
- Coles SG, Sparks RSJ (2006) Extreme value methods for modelling historical series of large volcanic magnitudes. In: Mader HM, Coles SG, Connor CB, Connor LJ (eds) *Statistics in Volcanology*, vol 1. Geological Society of London, Special Publication of IAVCEI, London, pp 47–56
- Deligne NI, Coles SG, Sparks RSJ (2010) Recurrence rates of large explosive volcanic eruptions. *J Geophys Res* 115:B06203
- Dilley M, Chen RS, Deichmann U, Lerner-Lam AL, Arnold M, Agwe J, Buys P, Kjekstad O, Lyon B, Yetman G (2005) Natural disaster hotspots: a global risk analysis. The International Bank for Reconstruction and Development, the World Bank, and Columbia University, Washington D.C
- Donovan A, Oppenheimer C, Bravo M (2012) Science at the policy interface: volcano-monitoring technologies and volcanic hazard management. *Bull Volcanol* 74:1005–1022
- Ewert JW, Harpel GJ (2004) In harm's way: population and volcanic risk. *Geotimes* 49:14–17
- Ewert JW, Swanson DA (1992) Monitoring volcanoes: techniques and strategies used by the staff of the Cascades Volcano Observatory, 1980 – 1990. *USGS Bulletin* 1966:223
- Furlan C (2010) Extreme value methods for modelling historical series of large volcanic magnitudes. *Stat Model* 10:113–132
- Guha-Sapir D, Below R (2006) Collecting data on disasters: easier said than done. *Asian Disaster Management News* 12:9–10
- Hittelman AM, Lockridge PA, Whiteside LS, Lander JF (2001) Interpretive pitfalls in historical hazards data. *Nat Hazard* 23:315–338
- IAVCEI (1951–1975) *Catalog of active volcanoes of the world and solfarata fields*. International Association of Volcanology and Chemistry of the Earth's Interior (22 volumes), Rome
- Lacroix A (1904) *La Montagne Pelée et ses éruptions*. Masson et Cie, Paris
- McEvedy C, Jones R (1978) *Atlas of world population history*. Penguin, London
- Newhall CG (2000) Mount St. Helens, master teacher. *Science* 288:1181–1183
- Newhall CG, Punongbayan RS (1996) *Fire and Mud: Eruptions and Lahars of Mount Pinatubo, Philippines*. Philippines Institute of Volcanology and Seismology, Quezon City, Philippines, and University of Washington Press, Seattle
- Newhall CG, Self S (1982) The volcanic explosivity index (VEI): an estimate of explosive magnitude for historical eruptions. *Journal of Geophysical Research (Oceans and Atmospheres)* 87:1231–1238
- Peterson DW (1986) Volcanoes: tectonic setting and impact on society. In: Geophysics Study Committee. *Active tectonics: impact on society*. National Academy Press, Washington DC
- Population Division, Population Estimates, and Projections Section of the United Nations Department of Economic and Social Affairs (2010) *World population prospects: the 2010 revision*. Available online: http://esa.un.org/unpd/wpp/unpp/panel_population.htm. Accessed on 1st and 2nd February 2012
- Populstat (1996/2006) *Population statistics: historical demography of all countries, their divisions, and towns (Populstat)*. Available online: <http://www.populstat.info>. Accessed on 30th and 31st January and 1st February 2012
- Siebert L, Simkin T (2002) *Volcanoes of the World: an Illustrated Catalog of Holocene Volcanoes and their Eruptions*. Smithsonian Institution, Global Volcanism Program Digital Information Series, GVP-3. Available online: <http://www.volcano.si.edu/gvp/world/>
- Siebert L, Simkin T, Kimberly P (2010) *Volcanoes of the World*, 3rd edn. University of California Press, Berkeley
- Simkin TA (1993) Terrestrial volcanism in space and time. *Annual Review of Earth and Planetary Sciences* 21:427–452
- Simkin TA, Siebert L, Blong R (2001) Volcano fatalities – lessons from the historical record. *Science* 291(5502):255
- Small C, Naumann T (2001) The global distribution of human population and recent volcanism. *Environmental Hazards* 3:93–109
- Sparks RSJ, Biggs J, Neubeck JW (2012) Monitoring volcanoes. *Science* 335:1310–1311
- Surono JP, Pallister J, Boichu M, Buongiorno MF, Budisantoso A, Costa F, Andreastuti S, Prata F, Schneider D, Clarisse L, Humaida H, Sumarti S, Bignami C, Griswold J, Carn S, Oppenheimer C, Lavigne F (2012) The 2010 explosive eruption of Java's Merapi volcano – a '100-year' event. *J Volcanol Geotherm Res* 241–242:121–135
- Tanguy JC, Ribièrre C, Scarth A, Tjetjep WS (1998) Victims from volcanic eruptions: a revised database. *Bull Volcanol* 60:137–144
- Tilling RI, Lipman PW (1993) Lessons in reducing volcano risk. *Nature* 364:277–279
- United Nations Development Programme (UNDP) (2011) *Human development report 2011*. United Nations Development Programme, New York
- United Nations Population Fund (UNFPA) (2011) *Population dynamics in the LDCs: challenges and opportunities for development and poverty reduction*. United Nations Population Fund, New York
- United Nations Statistics Division (2011) *Composition of geographical (continental) regions, geographical sub-regions, and selected economic and other groupings*. Available online: <http://unstats.un.org/unsd/methods/m49/m49regin.htm#developed>. Accessed on 10th February 2012
- Venzke E, Wunderman RW, McClelland L, Simkin T, Luhr JF, Siebert L, Sennert S, Mayberry G (2002) *Global Volcanism, 1968 to the Present*. Smithsonian Institution Global Volcanism Program Digital Information Series, GVP-4. Available online: <http://www.volcano.si.edu/reports/>
- Witham CS (2005) Volcanic disasters and incidents: a new database. *J Volcanol Geotherm Res* 148:191–233

doi:10.1186/2191-5040-2-2

Cite this article as: Auker et al.: A statistical analysis of the global historical volcanic fatalities record. *Journal of Applied Volcanology* 2013 2:2.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com