PAGER-CAT: A Composite Earthquake Catalog for Calibrating Global Fatality Models

Trevor I. Allen¹, Kristin D. Marano, Paul S. Earle, and David J. Wald
National Earthquake Information Center, U.S. Geological Survey

INTRODUCTION

The compilation of a comprehensive global earthquake catalog that delivers both accurate source parameters and fatality estimates is a task that is simple in theory but challenging in practice. The necessary information is spread throughout numerous earthquake catalogs, reports, and online databases. Earthquake catalogs are created for different purposes, and consequently they excel in different areas. Some catalogs provide high-quality hypocenters while others contain carefully researched damage reports. Herein we examine published global catalogs and create PAGER-CAT, a composite global catalog of earthquake source parameters and effects.

PAGER-CAT incorporates eight global earthquake catalogs and additional auxiliary data to provide comprehensive information not only for hypocentral locations, magnitudes, and human fatalities, but when available, focal mechanisms, the country of origin or the distance to the nearest landmass, local time and day of week, presence of secondary effects (e.g., tsunami, landslide, fire, or liquefaction) and deaths caused by these effects, the number of buildings damaged or destroyed, and the number of people injured or left homeless. The first version of the catalog is composed of more than 140 fields in which detailed event information can be recorded and currently includes events from 1900 through December 2007, with emphasis on earthquakes since 1973.

The catalog was compiled for calibration and development of earthquake fatality models to be used by the U.S. Geological Survey’s (USGS) Prompt Assessment of Global Earthquakes for Response (PAGER) system. The PAGER system currently provides estimates of the number of people and the names of cities exposed to severe shaking following significant earthquakes (Earle et al. 2008; Wald et al. 2008). In the future, PAGER will produce rapid fatality estimates within approximately 20 minutes of an earthquake’s occurrence anywhere on the globe, using loss models calibrated against PAGER-CAT (e.g., Jaiswal et al. 2008; Portet et al. 2008).

The development of PAGER fatality models from historical earthquakes requires estimates of the spatial variation of shaking intensity for several thousand global earthquakes. These estimates are contained in an Atlas of ShakeMaps (Allen et al. 2008) that was produced using the ShakeMap methodology (Wald et al. 1999). Generation of this atlas and subsequent PAGER fatality models required accurate earthquake source information and corresponding fatality estimates. Although unpublished, proprietary earthquake catalogs exist within the loss modeling community, we found no publicly available catalog containing both comprehensive earthquake source parameters and fatality information. This led us to develop a systematic approach to produce PAGER-CAT.

PAGER-CAT is available for download by the seismology and earthquake engineering communities at http://earthquake.usgs.gov/research/data/pager/. We hope the catalog and its associated products (e.g., the Atlas of ShakeMaps) will benefit other loss-estimation methodologies and risk analyses. All earthquake catalogs contain erroneous and incomplete data, and PAGER-CAT is no exception. Consequently, we encourage contributions from the community to further improve the catalog. Updates can be made directly to a condensed version of PAGER-CAT available on Wikipedia (http://en.wikipedia.org/wiki/List_of_deadly_earthquakes_since_1900). Alternatively, more detailed contributions and references can be forwarded to the authors for inclusion in subsequent catalog releases.

CONTRIBUTING EARTHQUAKE CATALOGS

PAGER-CAT is a composite earthquake catalog developed entirely from published or online databases and reports. No new information has been derived in its compilation. The catalogs chosen to develop the first version of PAGER-CAT include reliable, systematically generated global earthquake catalogs available in consistent digital formats.

We used eight global earthquake catalogs, compiled by recognized data centers and independent researchers. Each of the catalogs offered unique information regarding source and loss data. The catalogs were the USGS Preliminary Determination of Epicenters (PDE; Sipkin et al. 2000), the Centennial Catalog (Engdahl and Villasenor 2002; Villasenor and Engdahl 2007), the Engdahl, van der Hilst, and Buland (EHB) catalog (Engdahl et al. 1998), the Global Centroid Moment Tensor (GCMT) project catalog (Ekström et al. 2005; Global CMT Project 2006), the National Geophysical Data Center’s (NGDC) Significant Earthquake Database (Dunbar et al. 1992), the Utsu catalog of deadly earthquakes (Utsu 2002), the Centre for

¹. Contracted through Synergetics Inc.; now at Geoscience Australia

doi: 10.1785/gssrl.80.1.57
Research on the Epidemiology of Disasters' Emergency Events Database (EM-DAT; Hoyois et al. 2007), and the NGDC’s Historical Tsunami Database (HTD; Dunbar 2007). A more thorough review of these catalogs and other key global catalogs not directly used in the initial version of PAGER-CAT is found under the “supplementary information” link on the main PAGER-CAT Web page.

PARSING CASUALTY AND LOSS INFORMATION FROM THE PDE

The primary source for earthquake impact information for PAGER-CAT and several other global catalogs is the PDE catalog. The PDE contains a wide range of descriptive information regarding earthquake impacts. For significant global events since 1977, this catalog is available online at http://earthquake.usgs.gov/eqcenter/eqarchives/significant/. Unfortunately, much of this information is only available as human-readable comments that accompany an earthquake’s origin, which cannot be easily queried. Consequently, we parsed this information into machine-readable fields that are more convenient and useful to researchers.

Great care was taken to compile computer-readable fields from the PDE comments. For each significant earthquake comment, we extract information contained in the PDE that may include the number of fatalities and injuries caused by shaking (defined hereafter as deaths caused by partial or total structural collapse) and other secondary effects, the number of buildings destroyed, and the number of people left homeless. We differentiate between causes of death where information regarding secondary earthquake effects is available. Because fatality numbers are often not well-known even months after an earthquake, the PDE comment generally contains an approximate number with some descriptive text. For the purposes of PAGER-CAT, we use the following rules for assigning our preferred PDE fatality numbers:

- At least $x$ deaths = $x$ deaths
- Between $x$ and $y$ deaths = mean ($x, y$) deaths
- Estimated, or unconfirmed $x$ deaths = $x$ deaths
- $x$ deaths in place $a$, and at least/estimated/etc. $y$ deaths in place $b$ = $x$ + $y$ deaths

Where we assign the number of fatalities, we provide a numerical flag that indicates the rule used to derive the final number and the quality of the estimate. Flag definitions are provided in the README file that accompanies the catalog download.

We recognize that some information is difficult to consistently document in global earthquake catalogs. For example, the number of people injured from an earthquake can be a very subjective measure depending on a local authority’s definition of an injury. Consequently, we advise caution when using data fields such as injuries or buildings damaged/destroyed because of the inherent uncertainty involved in categorizing and counting them in a uniform manner. Nevertheless, we believe that this data is still important for inclusion in the PAGER-CAT, because in many cases it is the only documentation of such information we found.

EVENT SELECTION

We use two catalogs for our primary event selection: the updated Centennial Catalog prior to 1973 (A. Villaseñor, written communication 2008) and the PDE thereafter. Prior to 1973, the Centennial Catalog includes almost 4,700 recorded earthquakes of $M_\text{L} \geq 5.5$ and greater (hereafter, $M_\text{L}$ is referred to as the preferred magnitude of the source catalog). In comparison, the full PDE contains entries for more than 570,000 earthquakes of all magnitudes from January 1973 through December 2007. We impose magnitude criteria to keep the catalog to a manageable number of events. PAGER-CAT includes all events in the Centennial Catalog prior to 1973 and all earthquakes with a PDE preferred magnitude of $M_\text{L} \geq 5.5$ or greater. In addition, PDE events located within stable continental regions with $M \geq 4.5$ were also included to aid ShakeMap hazard calibration studies (e.g., Allen et al. 2008). Stable continental regions were defined by the polygons digitized from Johnston et al. (1994).

All earthquakes in the PDE with reported human casualties (fatalities or injuries) were also included in PAGER-CAT, regardless of magnitude (excluding mining-triggered events). Events listed in the PDE from non-tectonic sources (e.g., explosions) were removed from our final catalog. These criteria limited PAGER-CAT to just over 22,000 earthquakes from 1900 through December 2007.

ASSIGNING EARTHQUAKE SOURCE AND IMPACT ATTRIBUTES IN PAGER-CAT

Using the Centennial Catalog as the primary event list prior to 1973 and the PDE thereafter, we associate our preferred hypocentral locations and magnitudes to PAGER-CAT using the logic described in our online documentation. In summary, where possible we use the Centennial Catalog as our preferred origin source because it uses more sophisticated location algorithms (Engdahl et al. 1998) than the PDE, particularly for older events in the catalog. We also use moment magnitude from the GCMT (formerly Harvard CMT) as our preferred magnitude. Since 1976, 86.2% of the events in the PAGER-CAT ($M \geq 5.0$) are associated with CMT solutions. Where these magnitude estimates are not available, we use the Centennial Catalog’s preferred magnitude prior to 1973 and the PDE preferred magnitude thereafter.

Human and structural impacts are subsequently associated, in addition to secondary effect indicators from the parsed PDE comments, Utsu catalog of deadly events, and the NGDC, HTD, and EM-DAT databases. We also associate a GLIDE (GLobal IDEntifier) number (GLIDE 2008). GLIDE numbers provide a unique reference number for any multihazard disaster around the world and are intended to be used as a practical method of accessing comprehensive post-disaster information from a wide range of institutions and relief organizations.

Once the source and impact attributes have been assembled, we associate additional spatial information to the catalog in a geographic information system framework. First, each earthquake in the catalog is assigned an International Organization
for Standardization (ISO) two-letter country code (ISO 2007) and distance (in degrees) to the nearest country’s landmass.

Since the time of day may have a bearing on the human impact observed (e.g., Lomnitz 1970; Scawthorn et al. 1978; Coburn et al. 1992; Ramirez and Peek-Asa 2005), we also associate a local time correction factor to each earthquake. A list of countries that currently observe daylight savings was obtained from the website http://timeanddate.com, and events that occurred in these regions were flagged. However, it should be noted that the local time we provide is not adjusted for daylight savings.

Finally, tsunami occurrence and run-up information are acquired from the HTD. Also provided are centroid moment tensor solutions and style of faulting from the HRV or GCMT catalogs. The association of all event information in PAGER-CAT was largely automated but required manual intervention if two or more earthquakes shared similar time, location, and/or magnitude windows. The logic for deriving our preferred casualty numbers is described in the online documentation.

**PAGER-CAT OVERVIEW**

PAGER-CAT includes more than 22,000 earthquakes drawn from the combined Centennial and PDE catalogs for the period of 1900 through December 2007. In all, 146 fields of information are available in PAGER-CAT (see README file for details).

PAGER-CAT includes 1,373 earthquakes since 1900 where at least one fatality was reported. Unfortunately, some of these events do not have a numerical value indicating the number of deaths. This is often the case for aftershocks that caused fatalities additional to the mainshock and a total death tally was only provided for the mainshock. For other events where fatalities were reported, the number of dead is simply not available. There are 1,242 events in PAGER-CAT that have a numerical value of total fatalities. Of these, 1,041 events have shaking deaths (from partial or total structural collapse) explicitly defined in the catalog. The assignment of shaking deaths in PAGER-CAT was often an assumption based on the lack of information to support the occurrence of deaths from secondary sources (e.g., tsunami, landslide, or fire). The histograms in Figure 1 summarize the events in the catalog that have numerical values for shaking deaths, total deaths, number of injuries, and the number of people left homeless. Of significant interest in this figure is the number of homeless following a major earthquake (Figure 1D). These data strongly emphasize the need for sustained post-disaster recovery assistance following destructive earthquakes, particularly for less-developed regions of the world. As an example of the potential future risk to rapidly

![](image) **Figure 1.** Histograms indicating the “number of (A) shaking deaths (from partial or total building collapse), (B) total earthquake deaths, (C) the number of injuries, and (D) the number of homeless from global earthquakes. In each case, $N$ is the number of events since 1900 in PAGER-CAT with numerical values greater than zero.
Growing global communities, Figure 2 shows deadly historical earthquakes within PAGER-CAT with respect to the modern population distribution of the Asia-Pacific region.

**Comparison of PDE and Utsu Catalogs**

There are several catalogs that provide the number of earthquake fatalities. Those already mentioned herein include the PDE, NGDC, Utsu, and EM-DAT catalogs. Here we briefly compare fatality estimates between the PDE and Utsu catalogs. Use of these types of catalogs is fundamental to the development of global earthquake fatality models (e.g., Samardjieva and Badal 2002; Nichols and Beavers 2003; Badal et al. 2005).

We limit the period of comparison from September 1968 (when earthquake impact comments first appeared in the PDE) to 25 December 2003. We end the comparison period in late 2003 because the online Utsu catalog does not provide a numerical value for the 2003 Bam, Iran earthquake, and this may bias our comparisons. Table 1 summarizes some of the key differences between fatalities recorded in the PDE and Utsu catalogs, respectively. For the period of interest, the PDE catalog includes 75 earthquakes that caused fatalities that were not recorded in Utsu. These events represent a total of just over 800 fatalities for events where Utsu does not have any entries. However, many of these events have fewer than 10 recorded fatalities (Figure 3). By comparison, the Utsu catalog contains several events with death tolls in the range of hundreds to thousands for which fatality figures are not given in the PDE, summing to a total of over 21,000 fatalities for the same time period. One possible reason for the discrepancies is the aforementioned fact that the PDE catalog is not updated after publication. Utsu (2002) has the advantage of being more retrospective and was able to access historic data that may have been difficult to obtain when the monthly PDE was published. For example, the Utsu catalog includes fatality statistics from many deadly Chinese earthquakes that the PDE does not, particularly prior to 1980. It includes four Chinese earthquakes where greater than 1,000 fatalities were reported, the deadliest of which was a 1970 earthquake that killed more than 15,600 people.
TABLE 1
Comparison of earthquakes in the PDE and Utsu catalogs that indicate total fatalities, September 1968 to December 2003.

<table>
<thead>
<tr>
<th></th>
<th>PDE</th>
<th>Utsu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total events indicating deaths</td>
<td>558</td>
<td>541</td>
</tr>
<tr>
<td>Events unique to the catalog</td>
<td>75</td>
<td>58</td>
</tr>
<tr>
<td>Total missed deaths from events unique to the other catalog</td>
<td>21,160</td>
<td>818</td>
</tr>
<tr>
<td>Maximum deaths missed from single event unique to the other catalog</td>
<td>15,621</td>
<td>261</td>
</tr>
</tbody>
</table>

CATALOG DOWNLOADS AND EXTERNAL CONTRIBUTIONS

Periodically updated versions of PAGER-CAT will be available at http://earthquake.usgs.gov/research/data/pager/. The catalog is available for download as a comma-delimited text file, which can be imported into geographical information systems such as ESRI ArcGIS or as a Matlab data file that can be queried with an accompanying function.

To solicit external contributions and corrections to the catalog, we have made a partial version of PAGER-CAT available via the Wikipedia page “List of deadly earthquakes since 1900” (http://en.wikipedia.org/wiki/List_of_deadly_earthquakes_since_1900). The list combines fatality estimates from the PDE, Utsu, and EM-DAT, coupled with earthquake source information from the PDE, EHB, and GCMT catalogs. In the spirit of Wikipedia (e.g., Lih 2004), we add an additional column for third parties to contribute improved fatality estimates from official reports and detailed studies, where available. It is preferable that contributions be accompanied by a reliable citation. Verified contributions to the Wikipedia page would subsequently be reflected in future revisions of PAGER-CAT. We also encourage researchers to contact the authors directly for suggested modifications or more detailed contributions.

SUMMARY

We have described the compilation and contents of PAGER-CAT, an earthquake fatality model developed principally for calibrating earthquake fatality models. It brings together information from a range of sources in a comprehensive, easy to use digital format. Earthquake source information (e.g., origin time, hypocenter, and magnitude) contained in PAGER-CAT has been used to develop an Atlas of ShakeMaps of historical earthquakes (Allen et al. 2008) that can subsequently be used to estimate the population exposed to various levels of ground shaking (Wald et al. 2008). These measures will ultimately yield improved earthquake loss models employing the uniform hazard mapping methods of ShakeMap.

Currently PAGER-CAT does not consistently contain indicators of landslide and liquefaction occurrence prior to 1973. In future PAGER-CAT releases we plan to better document the incidence of these secondary hazards. This information is contained in some existing global catalogs but is far from complete and often difficult to parse. Landslide and liquefaction hazards can be important factors contributing to earthquake losses (e.g., Marano et al. unpublished). Consequently, the absence of secondary hazard indicators in PAGER-CAT, particularly for events prior to 1973, could be misleading to some users concerned with ground-shaking-related losses.

We have applied our best judgment in the selection of PAGER-CAT’s preferred source parameters and earthquake effects. We acknowledge the creation of a composite catalog always requires subjective decisions, but we believe PAGER-CAT represents a significant step forward in bringing together the best available estimates of earthquake source parameters and reports of earthquake effects. All information considered in PAGER-CAT is stored as provided in its native catalog so that other users can modify PAGER preferred parameters based on their specific needs or opinions.

As with all catalogs, the values of some parameters listed in PAGER-CAT are highly uncertain, particularly the casualty numbers, which must be regarded as estimates rather than firm numbers for many earthquakes. Consequently, we encourage contributions from the seismology and earthquake engineering communities to further improve this resource via the Wikipedia page and personal communications, for the benefit of the whole community.

ACKNOWLEDGMENTS

We would like to thank Antonio Villaseñor for his comments and suggestions, which helped improve PAGER-CAT and this manuscript. We also thank Jim Dewey for various discussions during the compilation of the catalog. Mike Hearne wrote the Matlab script to query PAGER-CAT. We thank Paula Dunbar for providing more detailed earthquake casualty information than available from standard Web queries on the NGDC catalogs and Lisa Wald for creating the associated Web page for PAGER-CAT downloads. Kishor Jaiswal provided a parsed version of the EM-DAT catalog. Thanks to Jim Dewey, Stuart Sipkin, Luciana Astiz, and an anonymous reviewer for
their reviews, which improved the manuscript. Finally, we are indebted to the compilers of the catalogs that form the basis of PAGER-CAT.

REFERENCES


Risk and Impact Analysis Group
Geoscience Australia
GPO Box 378
Canberra ACT 2601 Australia
trevor.allen@ga.gov.au
(T.A.)

National Earthquake Information Centre
U.S. Geological Survey
P.O. Box 25046, MS966
Denver, Colorado 80225-0046 U.S.A.
(K.D.M., D.J.W.)
pearle@usgs.gov
(P.S.E.)