# Annex A: the methodology behind IDMC's global estimates

IDMC's annual Global Estimates report aims to provide a broad, quantified, global view of displacement associated with disasters brought on by natural hazards, based on the highest quality data possible. Depth is addressed in a more anecdotal way, via case studies and other specific examples that are representative of underlying patterns. These notes provide further details of the methodology we use to produce our global estimates and the modelled trends introduced in section 1.1 and box 1.1 of the main body of the report. The annex is divided into two parts. The first part covers the annual global dataset of measured estimates for 2008 to 2013, and the second the modelling of historical displacement trends from 1970 to 2012.

## A.1 The annual measurement of displacement caused by disasters between 2008 and 2013

This section refers only to IDMC's annual global estimates based on direct reporting of displacement events. As explained in box 1.1, displacement is defined as the forced movement of individuals or groups of people from their homes or places of habitual residence, as described in the 1998 Guiding Principles on Internal Displacement. A rapid-onset shock in the form of a natural hazard may trigger such movements, as a result of its direct threat or impact on exposed and vulnerable people. The types of information used to monitor displacement include people reported as evacuated and people rendered homeless, as explained below.

This year's report presents the latest findings on displacement caused by disasters in 2013 and compares it with data on the six-year period from 2008 to 2013. We encountered regular challenges in the collection, compilation and interpretation of data from different sources, including varying institutional mandates, diverse research domains, differing terminology and definitions, and the variety of reasons organisations had for collecting and publishing the data in question. These are discussed further below.

### *Scope, resolution and limitations*

**Typological:** These estimates cover disasters associated with rapid-onset geophysical, climate and weather-related hazards, as shown in table A.1. Drought and gradual processes of environmental degradation are also significant drivers of disaster and displacement risk, but they are not covered in this report. They are excluded because a different methodology would be needed to analyse situations in which multiple stressors combine to create a point of crisis and displacement. Conceptual and methodological progress, particularly in relation to drought, is discussed in section 4.2.

**Spatial/geographical:** IDMC's data is monitored and collected with a broad global scope. We recorded displacement events induced by disasters in 161 countries over the six-year period from 2008 to 2013, and in 119 countries in 2013 alone. Event-based estimates can be aggregated to provide national, regional or global estimates, but the data does not allow for cross-event statistical analysis at the sub-national level. Nor is it currently possible to analyse the data by other location-related variables relevant to understanding exposure and vulnerability to hazards, such rural and urban settings, or mountainous, river basin and coastal areas.

We have increased our access to information at the country level over the past few years in a number of different ways: country missions by IDMC staff; cooperation with our colleagues in the Norwegian Refugee Council (NRC)'s country offices; and cooperation with other organisations such as IOM and IFRC that have country offices or national societies. Despite these efforts, our data compilation is still limited relative to the number of countries where displacement is known to have occurred. Our research is also limited by the working languages of our in-house experts, who work primarily in English, French and Spanish, and to a lesser extent in Italian, German, Russian and Japanese. That said, our access to local language sources has been improved through a partnership with IOM and its national and international staff.

For the purpose of this report, countries are defined as independent nation states. We do not analyse overseas territories or protectorates. For the few countries covered where sovereignty is contested - Kosovo/Serbia, Taiwan/China and Palestine - separate information was available and estimates were possible. The inclusion or exclusion of these and other contested territories does not imply any political endorsement or otherwise on IDMC's part.

**Temporal:** Data for each year since 2008 includes all identified displacements for which information was available from accepted sources as described below, and that started during the calendar year. It also includes a few events associated with disasters that started at the end of the previous year. In such cases, it was sometimes difficult to ascertain whether figures referred to displacement that began in the previous year or not, but we were careful to minimise the risk of double counting when a disaster, such as a period of flooding, ran from the end of one year into the beginning of the next. Consideration of a range of reports that described the disaster context as well as providing figures was helpful in this sense.

The estimates for each event or disaster represent the number of people reported as having become displaced at any point during them. They do not capture rates of return, the duration of displacement, the pattern of IDPs' movements after their initial flight or people living in prolonged displacement from one year to the next. For the time being, we are only able to report on repeated and complex movements and protracted and unresolved situations anecdotally or via case studies. This represents an important gap in terms of identifying displaced populations likely to be at particular risk and in need of protection and sustainable solutions.

**Demographic:** IDMC collects its data in ways that aim to be as inclusive as possible of all people displaced and without bias towards particular countries, segments of a population or in terms of where IDPs take refuge.

The displaced population in any given situation is far from a homogenous group, but access to disaggregated data is very limited. Analysis using key metrics such as gender and age is only possible for specific situations or seg-

*Table A.1* ***Typology of natural hazards***

<table>
<tr><th rowspan="2"></th><th rowspan="2">Geophysical</th><th>Meteorological</th><th>Hydrological</th><th>Climatological</th></tr>
<tr><th colspan="3">Weather and climate-related</th></tr>
<tr><td rowspan="3">Events/ shocks (rapid-onset)</td><td>Earthquakes: ground shaking, fault ruptures, landslides, liquefaction, subsidence, tsunamis and flooding</td><td>Storms: tropical storms (cyclones, hurricanes and typhoons), extra-tropical/winter storms, local storms (tornadoes, blizzards and snow storms, sand storms, hail storms, lightning)</td><td>Floods: land-borne or riverine floods (caused by heavy rains, snow melt, and breaking of banks), sea-borne or coastal floods (caused by storm surges and breaking of levees), flash floods (caused by snow melt run-off, dam bursts and sudden water release)</td><td>Wildfires: brush, forest, grass and savannah</td></tr>
<tr><td>Volcanic eruptions: explosive or effusive, lava flows and mud flows, falling ash and projectiles, toxic gases, floods, landslides and local tsunamis</td><td></td><td>Wet mass movements: landslides, avalanches and sudden subsidence</td><td>Extreme temperature: cold snaps and extreme winter conditions, heat waves</td></tr>
<tr><td>Dry mass movements: rock falls, landslides, avalanches, sudden subsidence and sink holes</td><td></td><td></td><td></td></tr>
<tr><td rowspan="2">Processes/ stressors (slow-onset)</td><td rowspan="2">Long-lasting subsidence</td><td rowspan="2"></td><td rowspan="2">Coastal erosion</td><td>Drought</td></tr>
<tr><td>Desertification</td></tr>
</table>

This table provides a non-exhaustive list of the types of hazards included in IDMC's displacement estimates and historical trend model. They are those loosely classified as rapid-onset events, shocks or triggers of displacement. This list also mentions some of those hazards not included, in particular drought. Specific hazards are often part of a series of sub-events that may take place over hours or months as part of a disaster, such as aftershocks and other secondary hazards that follow a major earthquake, or floods and landslides during or after a period of heavy rainfall. Classification for the purpose of this report refers to the original or primary hazard that triggered the disaster and displacement.

Categories are based on the classification system used by the International Disaster Database (EM-DAT), maintained by the Centre for Research on the Epidemiology of Disasters (CRED) in Brussels.

ments of the overall displaced population. Were such data more readily available, it would enable the statistical analysis of patterns and trends in the differentiated needs of people within and across diverse displacement contexts.

Higher quality data is usually limited to IDPs living in collective sites or settings, where they are assessed in more detail for operational purposes. Data on displaced people in dispersed situations outside official camps or collective sites is another important gap. This limits the ability of governments, humanitarian and development organisations and donors to prioritise where assistance is most needed.

For the purpose of this report, greater weight is given to providing as comprehensive an estimate of new displacement as possible, including IDPs living with host communities and in other dispersed settings, both in and outside the areas affected by a given disaster. As a result, the overall estimate for an event will be based on broader but less granular information sources if they are available. In many cases, however, the only information we are able to identify refers to a particular segment of the displaced population, such as those living in officially recognised collective sites, and the displacement figure we record is likely to be an underestimate.

### *Event-specific data*

IDMC only records new incidences of displacement in its annual datasets when the information available allows event-specific estimates to be made. We do not use figures that we are unable to break down because they are reported already aggregated at the national level, for a whole year or by type of disaster. This ensures consistency and comparability across the data captured. In a few cases, we were unable to incorporate official aggregated statistics made available to us into the dataset. It is worth noting, however, that in all of these cases the official statistics gave a higher estimate of displacement for the country or type of disaster than our own, probably because some events were missing from our data and/or because we underestimated the displacement involved in one or more of that year's events.

### *Defining a displacement event*

The data behind the annual global estimates and longer-term modelled trends include displacements of all sizes, ranging from a few records of only one person being displaced to mass displacements of more than 15 million people. The data compiled for 2013 includes more than 600 events, of which 375 displaced at least 100 people. Thirty-seven involved the displacement of between 100,000 and a million people, and there were six mega-scale events in which more than a million people were displaced. There were 34 mega-scale events between 2008 and 2013.

We compile our data without any lower threshold on the size of the displacement recorded. Where necessary, we used a threshold in our analysis to eliminate any bias caused by the irregular reporting of small events, by excluding those that displaced fewer than 100 people. The data sources available and our methodology create a bias towards larger, more visible and more widely reported events. Lesser disasters that cause frequent small-scale displacements are included for countries where this type of detailed information is available, such as Indonesia. From these, we can infer that small-scale events are significantly under-reported for most countries, as discussed in section 2, box 2.1.

Classifying and defining a disaster event period associated with displacement can be challenging, given that it may be difficult to determine its start and end date, its geographical scope and its complexity beyond the direct and initial impact of the hazard. In reality, a disaster usually involves a number of sub-events and phases. This is particularly true of displacement across wide areas during successive periods of heavy rain together with secondary impacts such as landslides, or when similar events happen in parallel or close succession in the same country or locality. As the Dartmouth Flood Observatory notes: "Repeat flooding in some regions is a complex phenomenon and may require a compromise between aggregating and dividing such events."[6] This issue does not change the overall number of people estimated as displaced, but it does affect the number of events recorded and analysis of those events according to their size.

The 2013 data includes a significant increase in the recording of smaller-scale extensive disasters. Highly detailed information on a large number of small local events was aggregated when they were clearly identifiable as related to a main weather system, flood season or other hazard, including secondary hazards such as landslides during a period of flooding. This type of aggregation is often used in the international reporting of disasters, and we applied it to 40 disasters in 10 countries. Better data provision and storage means that detailed records of the sub-events are maintained on our database, to facilitate more granular analysis in the future.

Our data also includes reported disasters for which no displacement was recorded. If information was not available to compile an estimate in accordance with our methodology, this was recorded as "no data available", while events for which sources explicitly stated that no displacement occurred were recorded as "zero displaced". The difference is important to note, because it is much more common for the scale of displacement associated with an event to be unknown than confirmed as zero.

In some cases, people fleeing a natural hazard or disaster were already living in displacement before it struck. If it was clear, for example, that people already displaced by conflict were then forced to flee again by an event such as the flooding of a displacement camp, their new movements were recorded as 2013 displacements caused by a natural hazard. It should be noted that only limited information is available on such displacements, meaning that the dataset captures only a small number of them and they are likely to be under-represented.

### *Sources of information*

IDMC regularly reviews the various types of information released by different sources on the number, needs and characteristics of displaced people, primarily by gathering and monitoring secondary reports. We systematically seek a range of sources for each country and each disaster. For our 2013 estimates, we increased our research capacity and accessed data from sources including the Asian Disaster Reduction Centre (ADRC)'s GLIDE website, IFRC's disaster management information system, OCHA and other UN agencies, IOM, humanitarian cluster situation reports, government reports and national disaster loss databases, and NGO reports. Reputable media sources provide citations of government officials and local authorities in affected countries. For small events, local media reports are often the only source of information available. IOM country offices provided field data and/or gave us access to official sources that we incorporated into our data for 28 countries.

### *Selection and calculation of estimates by event or disaster*

For the purpose of providing global estimates, IDMC aims to arrive at the best approximation of the total number of people displaced by a specific event or disaster, measuring the incidence of displacement rather than the evolution of the number of people displaced and their movements and situations over time. Our analysis and interpretation of information from multiple sources includes the cross-checking of reported locations and dates to ensure that figures are associated with the same disaster and time period, and that double counting is avoided or minimised. All new incidences of displacement during a given event or disaster period are recorded, which requires the analysis of reporting dates and the consideration of series of situation reports.

The estimate per event is selected according to the most accurate and reliable figure provided or calculated based on a single source, or combined sources when it is clear that overlap and double counting can be avoided. The number of sources available varies according to the scale of the event, from one or two for smaller events to more than four for larger events, disregarding those that republish original information from elsewhere. Disasters widely covered by media, or which continue for long periods of time, also tend to have more sources from which to draw.

A wide range of terms - such as evacuated, homeless, damaged and destroyed housing, fled, relocated and affected - definitions and methods are used for collecting and reporting figures, and they are used in different ways by different sources. Such variations arise in part from the different purposes organisations have for collecting and reporting their data in the first place. In operational settings, the term "displaced" is often applied more narrowly than IDMC's definition. It may be used to indicate only those people staying in official collective sites or camps, or only people displaced a certain distance from their homes.

In some operational contexts, evacuees who move to official, short-term evacuation centres are counted separately from displaced people in official camp-like shelters. In others, evacuees are counted as a subset of the displaced population. Displaced people are sometimes counted as a sub-set of the affected population, and sometimes as additional to them. Information describing the context and point in time at which displacement is reported, knowledge of typical patterns observed in similar contexts and the quality and reliability of different sources are also taken into account.

We interpret the data we collect using the same broad and inclusive definition of displaced people across all events worldwide. Our definition assumes that displaced people are part of the population affected by a disaster, though this does not imply that those affected have necessarily been displaced.[2] We consider evacuees to be displaced people whether or not their evacuation was pre-emptive (see box 1.1), and we define people whose homes are rendered uninhabitable as displaced, regardless of where they are displaced to, how near or far from their homes they move and whether they are able or not to return.

We recognise that different situations create different types of needs, but our research indicates that being displaced further away does not necessarily imply greater needs or vulnerability. Indeed, displacement over short distances, especially when recurrent, may be a better indicator of vulnerability, given that those affected may face limitations on their movement to safer locations or places where they have better access to assistance.

**Evacuation data:** In addition to people directly reported as displaced, having fled or been forced to leave their homes, one of the most common types of data used to estimate event-based displacement comes from mandatory evacuation reports and official evacuation centres. The number of people reported as staying in evacuation centres may underestimate the total number of evacuees, given that some may take refuge at unofficial sites or with family and friends. On the other hand, the number of people ordered to evacuate may overstate the true number of evacuees, given that some will usually not heed the order and remain in their homes. The potential for such discrepancies is much greater when authorities advise rather than order evacuation, and as a result we do not incorporate such figures into our estimates.

**Data on people made homeless and uninhabitable housing:** People made homeless because a disaster renders their homes or habitual residences uninhabitable are considered displaced. Their number may be reported directly, or we infer it based on the number of homes reported as severely damaged or destroyed, multiplied by the average household size for the country in question. In the absence of international and standardised average national household size information for all countries, we apply a consistent calculation by assuming two adults per household plus the total national fertility rate, as provided by UN Statistics for 2010 to 2015. We do not use data on housing reported simply as damaged, because the term is too broad to determine whether it has been made uninhabitable or not unless the source itself makes the fact clear.

Data on homelessness also points to the risk of prolonged displacement and the severity of the situation. Areas where homes and community infrastructure have been severely damaged or destroyed are unlikely to be able to support early safe returns. Migration from rural to urban areas, a lack of social housing for poor families, the unplanned growth of informal settlements and the failure to implement building standards for disaster-resilient housing puts millions of people at risk of being made homeless, with the poorest being the most vulnerable.

### *Reporting bias*

Given the issues discussed above, IDMC's overall annual estimates are likely to underestimate the scale of displacement around the world each year. There are a number of causes of bias that should be noted, both in our source information and our methodology:

- It is often difficult to determine whether displacement data for many countries and situations is reliable and comprehensive. Global reporting tends to emphasise large events in a small number of countries where international agencies, donors and media have a substantial presence, or where there is a strong national commitment to, and capacity for, disaster risk and information management. Information on smaller-scale disasters is far scarcer and on the whole significantly under-represented. The effects of disasters on isolated and insecure areas also tend to go relatively unreported because access and communications are limited.
- There tends to be significantly more information available on displaced people in official or managed collective sites than there is on those living with host families and communities or in other dispersed settings. Given that the majority of IDPs usually fall into the second category (see section 2.1.3), figures based on data for collective sites only are likely to be substantial underestimates.
- Reporting tends to be more frequent but also less reliable in the most acute and highly dynamic phases of a disaster, when peak levels of displacement are likely to be reached. It becomes more accurate once there has been time to make more reliable assessments. This means that estimates based on later evaluations of severely damaged or destroyed housing will be more reliable, but they are also likely to understate the peak level of displacement, given that they will not include people whose homes escaped severe damage but who fled for other reasons.
- It should also be noted that reporting bodies may have interests in manipulating the number of people displaced. They may be to maximise the potential for receiving external assistance, downplay the scale of a disaster if the government may be held accountable, or because international attention is deemed politically undesirable.
- A time delay in the updating of national and international disaster loss databases means that some information was not available in the research period for this report. The 2013 dataset did, however, include information from more national disaster loss databases than in previous years because it has been scheduled for publication a few months later.

Improvements in the systematic collection and sharing of reliable information on displacement are essential if we are to continue to improve the quality of our reporting and monitoring - a critical first step in identifying needs, prioritising assistance and informing longer-term solutions.

### *Data contributions and review*

Ahead of this year's report, IDMC's event-based datasets for 2008 to 2013 underwent significant improvements in terms of normalisation and standardisation, which has increased the type and quality of analytics we can run. We also made a substantial effort to increase the background information we collected for displacement events. The process for screening all data and estimates has also been improved, including the introduction of more thorough checks by two additional in-house researchers, a focused review of all larger events, and the review of country data by IDMC analysts in consultation with in-country contacts and NRC country offices. IOM field and liaison offices around the world also provided extensive inputs, and our REACH Initiative partners reviewed data from a number of countries where they maintain a presence.

We also carried out a review of our previous data from 2008 to 2012 in order to better standardise our database. Our first Global Estimates report, for 2008, added figures for displaced people to those for evacuees, but our methodology for 2009 to 2013 works on the assumption that the two sets of figures often refer to the same people. As such, the review aimed to eliminate any double counting in our 2008 data and improve comparability across the last six years. For the 100 events that displaced more than 5,000 people, the sources were double-checked and additional research was carried out to verify our estimates. A small number of corrections and updates were also made to our estimates for 2011 and 2012, based on errors identified and new information brought to our attention.

To improve the quality and comprehensive nature of the data we use to produce our global estimates each year, we collaborate with an increasing number of partner organisations, and incorporate lessons learned from previous years. Feedback on our work and suggestions for the future are always very welcome.

## A.2 Modelling displacement trends from 1970 to 2012

IDMC expanded the evidence base for this year's Global Estimates report by including 43 years of historical data on disaster impacts, which we used to model displacement trends from 1970 to 2012. The main reason for doing so was to provide a broader historical context for displacement associated with disasters, because our dataset for 2008 to 2013 covered too short a period of time to detect and analyse long-term trends and patterns. The exercise also helped bring to light patterns and relationships between displacement and other variables. It made clear, for example, that there can be huge variations year to year in the number and type of hazards that occur, and in the amount of displacement they cause. Some hazards occur frequently and with relative predictability, perhaps numerous times a year in a particular country or region, while others may happen once every 500 or 1,000 years worldwide. This is an important consideration when one takes into account the natural human tendency to draw conclusions from available data. Our dataset for 2008 to 2013 is reasonably accurate, but it is not necessarily representative of any other six-year sample in the last century, nor is it alone a good basis on which to predict future displacement trends.

We undertook our trend-based analysis with several important caveats in mind. First, the sample sizes are too small given the available data to make inferences about individual countries. This applies particularly to small territories and populations, and those relatively unexposed to hazards that only occur rarely, both of which may only be recorded a few times if at all in either the six or 43-year datasets. Trends based on region, continent or other means of grouping countries with similar characteristics together are more likely to produce accurate and meaningful results. It is also important to note that the 1970 to 2012 trends were modelled and subsequently calibrated using datasets that overlap for only a five-year period from 2008 to 2012. We obtained the additional data for the analysis, covering 1970 to 2007, from the EM-DAT international disaster database[3], national disaster loss databases[4], and datasets from the World Bank, the UN and other demographic sources.

At first iteration, the modelled displacement estimates provided some validation of observed patterns. They also shed light on some interesting relationships between datasets and trends. Taken as whole, the exercise highlighted opportunities for future research and for the improvement of our analysis of displacement patterns, both in terms of underlying data and modelling methodology.

The expansion of our dataset to include close analysis of major historic events and successive years of displacement data has created a much larger sample size for future calibration efforts. Successive iterations of the model using improved and expanded data, together with ongoing refinements to the calculation methodology, should reduce uncertainty in future analyses and expand their descriptive and predictive capacities.

### *Datasets used for modelling displacement*

We used direct proxies for displacement in the creation of our dataset covering 2008 onwards, including evacuation figures, people living in temporary shelter sites and homes destroyed. Comparable direct, high-quality and consistently recorded proxies are not, however, readily available at the global level for the entire 1970 to 2012 period.

Given the limited availability of such data, we used direct proxies such as recorded homelessness data, or figures for people requiring shelter during a disaster, with indirect proxies such as the number of people affected and the number of people killed – some of the most common types of data collected for disasters on a historic basis. At the global level, EM-DAT is the most thorough and most often cited database of disaster impacts and losses that tracks these variables.[5] Data on the number of homes destroyed, for example, is a particularly good proxy for displacement in earthquake scenarios. Disaster-related mortality may, at first glance, seem an strange proxy for displacement, but statistical analysis shows that for certain hazards, such as floods, there is a correlation between the number of people killed and the number displaced.

At the national level, a growing number of countries have begun to develop disaster loss databases using the DesInventar methodology, which provides disaggregated and geospatially referenced data on a number of disaster impacts and variables. National DesInventar databases were first implemented in Latin America in the late 1990s to satisfy a need for disaggregated, local-level information on disaster losses in order to better understand patterns across territorial, political and economic zones. In many cases, the databases contain very detailed information across a wide range of categories according to each country's specific information needs. Given the nature of the data involved, efforts to build and maintain them have often involved civil society organisations, drawing on their extensive access to local-level information. Such linkages between civil society and government agencies foster better disaster risk management by bringing relevant stakeholders together, which in turn leads to improved outcomes at the local level.

As each country administers its own DesInventar database, there are slight variations in structure and more significant variations in data entry, coverage and thresholds - such as the number of deaths or people affected - which determine whether an event is included. Most, however, record information on a wide range of indicators, from damage to health facilities to secondary and downstream economic losses.

In both the EM-DAT and national databases, mortality data is of better quality than that on people affected or rendered homeless. The difference in quality also varies from hazard to hazard. Homelessness data, for example, appears to be most accurately represented for earthquakes, and least well-tracked for smaller floods. Disasters linked to storms and major floods have both the highest number of entries and largest total figures for people killed, affected or left homeless. Given the larger sample size available for these hazard types, subsequent results and analyses are generally more robust.

Disasters linked to frequently occurring and localised hazards such as landslides and small seasonal floods receive substantially less attention because of the difficulties in collecting data on so many events, and differences in methodology such as the thresholds used for inclusion. EM-DAT's threshold for including an event is 10 deaths or 100 people affected, which means that the data is likely to be biased towards events in which one or both of these metrics are met, and against events during which homes are damaged or destroyed and livelihoods lost or severely disrupted. This is just one example of a bias that makes analysis challenging. Similar variability occurs across hazards and loss metrics as well as databases.

### *Modelling and calibration with the 2008 to 2013 dataset*

Our 1970 to 2012 model was calibrated using our high-quality 2008 to 2013 dataset. Country, hazard type and annual data from both were compared. It is important to note that the overlapping years between the datasets provide only a limited sample, which may not be representative of disaster impacts and displacement over the 1970 to 2012 period. We will address this limitation by continuing to research additional years and past events.

Three iterations of the process were run, seeking to improve the predictive capabilities and reduce sources of uncertainty in the results generated by the historical model. The procedure benefitted substantially from the improvements to the 2008 to 2013 dataset mentioned above, and both employ a similar data structure, extending analytical capacities and enabling direct comparison between them.

The first iteration, based on EM-DAT disaster loss data, applied a "naïve" multiplier across all hazard types. This had the benefit of providing a rough estimate without any significant variance issues, but it failed to produce a good fit in terms of underlying hazard, country and annual data when compared with the events in our existing dataset for 2008 to 2012.

The second iteration used regression coefficients for each hazard, where possible, and generic values for hazard types with limited samples. This meant that the impacts of different hazards were weighted more realistically. The third iteration sought to address some of the challenges the second model raised by using relative values and increasing the sample size of disaster events.

The second and third iterations were calibrated using coefficients obtained from regression analyses between our annual displacement totals by country and year for 2008 to 2012, and equivalent annual mortality, affected and homeless data by country from EM-DAT. For most hazard types, the regressions were run with data corresponding to each one. For hazard types with limited data - landslides triggered by earthquakes, for example - values were obtained from regression analysis across all the hazard types we identify.

Given the limited sample sizes, the divergence of exogenous variables over the 1970 to 2012 period was much larger than in the 2008 to 2012 sample used for the regressions. As a result, some entries appeared as extreme outliers, thereby skewing the results. Several approaches were taken to deal with the most extreme outliers generated in the second iteration of the model, including scaling values to mortality, affected, homeless and displaced figures expressed per million inhabitants.

### *Preliminary results*

The third iteration of the model provides some validation of expected displacement patterns, and analysis of the results identified new avenues for future research. The initial results presented here are compared with global trends in the key metrics on which the model is based, for example people affected by disasters, people killed by the same type of disaster and people reported as left homeless by disasters over the same period of time.[6]

The overall model shows a clear upward trend over the past four decades, with displacement associated with all types of weather-related hazards and earthquakes increasing. The model also suggests that the average number of people displaced each year has doubled since 1970, with displacement associated with floods rising fastest (see figure A.2). The trend mirrors the increase in the number of people affected by the events in question, which is assumed to include people displaced as per IDMC's definition (see figure A.1a).

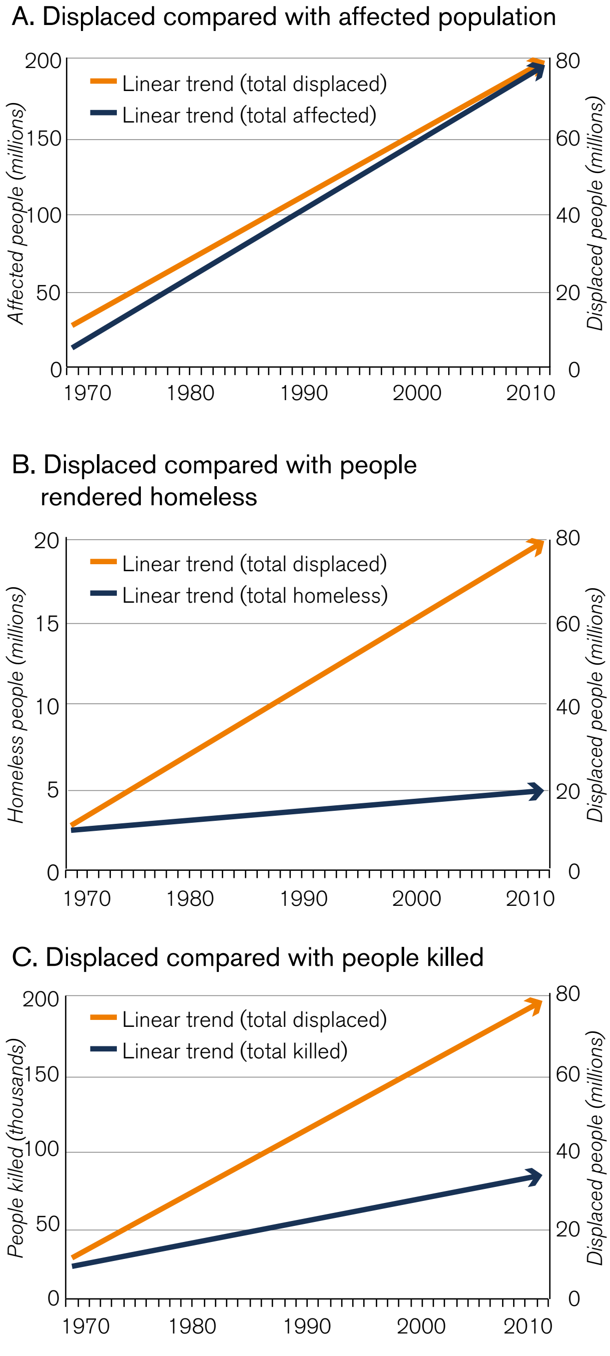
The number of people rendered homeless by disasters has also increased since 1970, albeit at a slower rate than those displaced. Not all people who are displaced have necessarily been left homeless or in need of shelter assistance. They may have fled pre-emptively to avoid a potential threat to their safety, or because they no longer had the access to land, livelihoods and services required to meet their basic needs (see figure A.1b).

The number of people affected by disasters rose significantly over the 43-year period, but absolute mortality related to all types of hazards combined increased only slightly,[7] pointing to improvements in preparedness, early warning systems and other life-saving measures (see figure A.1c). If the trend continues, disaster-related mortality may become a weaker proxy for displacement. If there are fewer fatalities relative to the number of people affected, it is likely to mean that more people are being displaced.

The affected, homeless and mortality metrics underlying the model exhibit different patterns over the 43-year sample period. The average number of recorded disasters and people affected both increased substantially, but the consensus points to a relatively flat trend in the number of people killed, and a decreasing trend in the number of people killed relative to the total population and those exposed to hazards. Increases in other categories are commonly attributed to a combination of greater exposure and improved reporting.

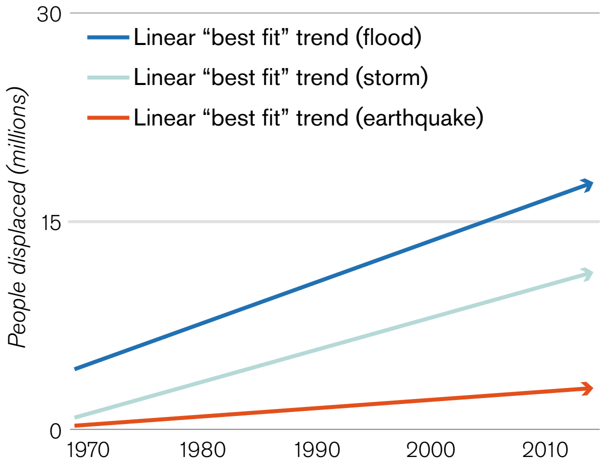
The metrics also exhibit different patterns based on hazard type and other exposure-related variables. The relationship between a rising displacement trend and an almost flat mortality trend, for example, can be seen most clearly in the modelled results for storms. There are likely to be a number of reasons for this. Storm-related deaths may have fallen thanks to improved weather forecasting, early warning systems, pre-emptive evacuations and disaster preparedness. At the same time, rapid, poorly planned and unregulated development means that increasing numbers of people are exposed to hazards and may be forced to flee their homes.

***Figure A.1, a-c: Modelled historical displacement trend compared with people affected, rendered homeless and killed by disasters, 1970-2012***



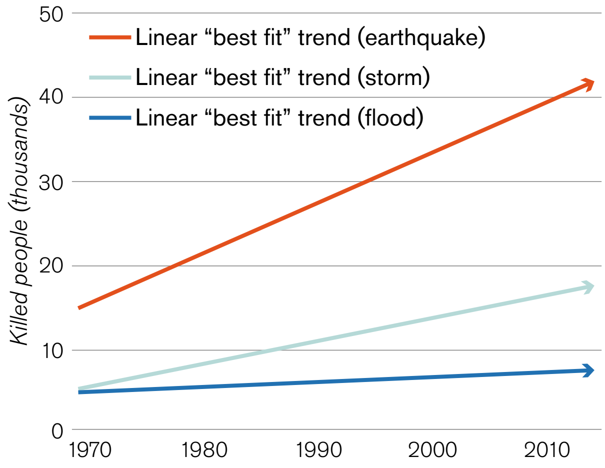

Source: IDMC-modeled trends as of 22 August 2014

*Figure A.2* ***Modeled displacement trends by hazard type, 1970-2012***




Source: IDMC-modeled trends as of 22 August 2014

*Figure A.3* ***Disaster mortality trends by hazard type, 1970 - 2012***




Source: IDMC-modeled trends as of 22 August 2014

Flood mortality data correlates very closely with modelled displacement. This is indicative both of fewer flood-related deaths being recorded relative to other types of disaster, and of the way the model weighted flood mortality to estimate historic displacement (see figure A.3). An early conjecture on this relationship could be that small, localised floods are relatively easy to escape, while larger more widespread floods that trigger major displacements pose a higher risk of mortality and entail greater risks associated with mass movements of people, some of whom are unable to flee to safety. More exhaustive causal analysis and significantly more and better data is needed to identify and understand such underlying relationships.

Earthquake data shows a rising trend for both displacement and mortality, indicative of the lack of advance warning such events allow for, and the importance of long-term risk reduction measures such as building and zoning regulations that reduce exposure and vulnerability in high-risk areas. It also demonstrates the highest correlation between modelled displacement and homelessness figures, and as such is weighted by the regression model coefficients to rely more heavily on this data than the other hazard types.

There is relatively little data on volcanic eruptions, landslides, wildfires and extreme temperature events with which to compare the 1970 to 2012 and 2008 to 2013 datasets. As a result, the modelled displacement estimates generated for these hazards must be interpreted with particular caution and are not included in the graphs above (see figures A.2 and A.3).

### *Next steps*

Several potential areas of improvement have been identified for the next iteration of both the 2008 to 2013 and 1970 to 2012 datasets. Further investigation into causal relationships between underlying risk drivers and displacement is also envisaged, stemming from the increased analytical capacities of both datasets. This includes comparison with demographic, social, economic, land-use, governance and other variables. Ongoing improvements in data management, review tracking and source document archiving continue to improve the depth and breadth of the datasets.

An expansion of event-by-event coverage to include prominently large displacements over recent decades, and to focus on hazard types for which sample sizes are highly limited, will help increase the robustness of the calibration algorithm. Event-by-event matching between the 2008 to 2013 and the historical data, at least for the top 50 per cent of entries, would also help address another limitation in the current model - the compilation of annual data by country, rather than event-by-event, means that some events that caused large disaster losses may skew the annual estimate for the country in question.

One last limitation to note is that disaster loss data was compiled on an annual basis, both to keep the size of the dataset manageable and, more importantly, to enable matching by year, hazard and country between the two datasets. Otherwise calibration of the historic loss data would be impossible. To address the issue, more event-by-event matching between the two datasets is planned to further validate and expand on the data and findings to date.