METHODOLOGICAL ANNEX

Introduction

The figures included in this report are the result of IDMC's most ambitious effort yet to present our estimates as transparently as possible. We have also attempted to apply more methodological consistency to our data collection and analysis, and to document this process for our readers. These improvements have helped bring our reporting on displacement associated with disasters and that associated with conflict and violence together in one report. They have also enabled us to make more rigorous comparisons between different displacement situations and get more out of our source data.

The evidence presented represents a baseline, and indicates many areas in which we will need to improve our data gathering and analysis in order to paint a comprehensive picture of internal displacement. This section highlights some of the main challenges we face and illustrates the most significant caveats to which we call readers' attention.

Our data on displacement associated with disasters for 2016 covers 591 displacement events triggered by sudden-onset natural hazards in 118 countries and territories. We are still in the process of developing and extending our approach to monitoring displacement associated with drought and other slow-onset phenomena, which means we do not yet have global figures for such disasters (see part 3).

Our data on displacement associated with conflict and violence covers 55 countries and one disputed territory. We have data on several other countries, but we chose not to include it in our global figures for methodological consistency.

As we did last year, as part of our innovative methodology we are also providing our assessment of confidence in the primary data and what it means for the estimates concerned. The confidence assessments signal our commitment to transparency while providing a roadmap for future work to strengthen data collection, something we are committed to helping our partners achieve over the coming years.

This annex describes how we produce our displacement figures by explaining the source data, calculations, definitions and decision rules we use in our analysis. Our aim is to provide maximum transparency so that readers understand the process, can replicate our work independently and make use of our data in innovative ways. We will make our data publicly available on our website for others to use freely.

We are also using innovative ways of allowing policymakers, researchers, partners, the media and the public to interact with our data via an open portal, making it easier to produce customised reports and analyses.

Given the complexity of displacement, we are forced to rely on a variety of internal and external sources in compiling our estimates. We have reassessed some of the criteria we use to maximise the reliability and accuracy of source data, and this report presents our figures in a way that clearly indicates how recently it was updated.

We currently use two similar but distinct methodologies to produce displacement estimates related to conflict and violence, and disasters. This annex describes both approaches.

To monitor and report on displacement associated with conflict and violence, we collect data on the countries affected and present nationally aggregated figures for:

| New incidents of displacement from 1 January to 31 December 2016

| IDPs who returned, integrated locally or settled elsewhere between the same dates, and when available, for those who crossed an international border and those who were born or died in displacement

| The total number of IDPs as of 31 December 2016

We use an event-based methodology to estimate the number of people displaced by disasters during the course of the year, and derive aggregated figures for new displacement for each of the countries affected.

We have monitored displacement associated with conflict and violence since 1998 and that associated with disasters since 2008. We have continuously sought to improve the ways we collect and analyse our data, and over the past nine years we have successfully obtained data on ever larger numbers of new displacement events associated with disasters, accounting for more small to medium-sized events than in previous years (see table A.1). Reporting on these events helps paint a more comprehensive picture in terms of the number of people displaced globally. It also provides an empirical evidence base with which to understand them and how they differ from mega-events.

Table A.1 Categories of events by magnitude

Event size	Number of people displaced
Small to medium	Fewer than 100,000
Large	100,000 to 999,999
Very large	One to three million
Mega	More than three million

As a result of ongoing methodological improvements, including the way partners collect data and the standardised application of the rules and criteria used to analyse displacement associated with conflict, comparisons between countries are now more valid than before.

Relating others' data to IDMC's data model

In order to obtain a comprehensive and accurate picture about the scope and scale of displacement at any given point in time, we have generated a unique data model (see figure A.1). One of the challenges we face in producing our figures is relating our partners' primary and secondary data to it.

In order to account comprehensively for the number of people displaced in a given situation, we would have to populate each component of the model, updating the information as quickly as the situation evolved. We are currently working with partners such as IOM, OCHA and UNHCR to do just that, in an effort to better reflect the dynamics of displacement.

The purpose of our data model is to better capture all incidents of new displacement, or "flows", during the year as information becomes available, the number of IDPs reported to have found durable solutions or to have crossed an international border, the number of children born in displacement and the number of IDPs who have died.

Figure A.1: IDMC's displacement data model
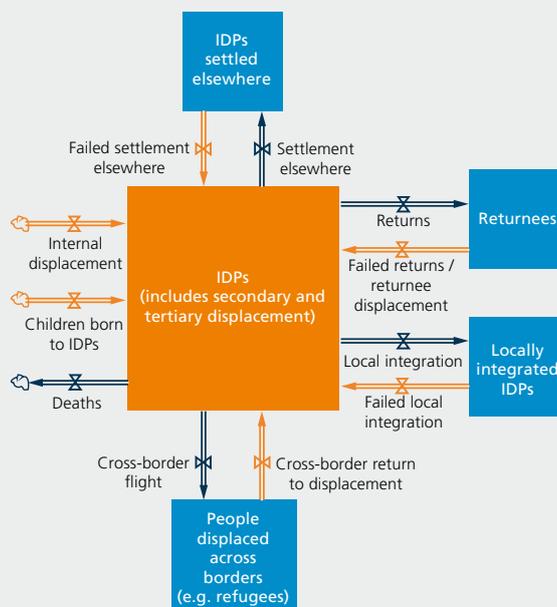
The model is an ideal vehicle for compiling displacement estimates, but in reality we have found it difficult to populate systematically. We seldom receive comprehensive data from our partners for all of its components. This is often because the type of data specified is simply not collected or, when it is collected, it is not disaggregated. A primary data source may report the extent to which the number of IDPs has declined during the course of the year, but may not specify the reason for the decrease.

The remainder of this annex explains how we account for the main flows we report, and how they influence our estimates. It also explains how we have selected countries and events to include and why we have excluded some countries we have reported on in the past. It also outlines how we assess and express our confidence in the source data.

We have continued to harmonise the approaches we use to monitor displacement associated with conflict and disasters – by identifying more events that caused displacement in the context of conflicts and by capturing more time-series data on caseloads of people displaced by disasters. That said, there are still some differences between the two approaches which reflect the availability of data and our ability to detect certain events and processes (see table A.2).

Standardising the data collection

Countries and contested territories

We use the ISO 3166-1 alpha-3 standard for coding countries and for mapping, but as the territories of Kosovo and Abyei do not have an official code assigned, we adopted the following: Kosovo (XKX) and Abyei (AB9).

The geographical referential we use is based on datasets such as the Global Administrative Areas (GADM) and the Global Administrative Unit Layers (GAUL) and other sources. The designations do not imply IDMC's official endorsement or acceptance.

Additional notes:

| The Kosovo designation is in line with UN Security Council resolution 1244/1999 and the International Court of Justice's opinion on Kosovo's declaration of independence.

| As the status of the Abyei area is not yet determined, for the purpose of monitoring we used the border representation of the 2005 peace agreement between the Sudanese government and the Sudan People's Liberation Movement.

Table A.2: Comparison of main monitoring attributes for displacement associated with conflict and disasters

Displacement monitoring attribute	Conflict and violence	Disasters
Event-based	Partial	Yes
Geography or situation-based	Yes	Partial
Global coverage	Yes	Yes
Quantitative threshold	No	No
Enables reporting of number, or stock of IDPs	Yes	Not at the global level
Covers incidents of new displacement	Yes	Yes
Includes other inflows and outflows that determine the number of IDPs	Yes, subject to availability	No, lack of data
Includes SADD	Yes, subject to availability	Yes, subject to availability
Figures disaggregated based on age of source data	Yes	Not applicable
Application of AHHS data	Yes	Yes

Population data

We use the 2015 UN World Population Prospect (WPP15) as our reference for population data.[1] The 2016 population estimates are based on the medium fertility variant projection.

Normalising displacement data by country population size

To illustrate the magnitude of internal displacement at the country level, we normalise the data to account for population size. In doing so, a clear distinction has to be made between the notion of population and inhabitants. When displacement is acute, including refugees fleeing across international borders, the population in a country at a given time may be significantly lower than the official figure.

Syria is the most graphic case in point, but the issue also affects other countries such as Libya and Somalia, for which there are no up-to-date and reliable national population figures. As such, the ratios of IDPs to population and inhabitants will differ, but both provide useful information for research and analysis.

Income groups and geographical region

Income groups and geographical groups are based on the World Bank's classification.[2]

Accounting for displacement associated with conflict and violence

We produce our figures for displacement associated with conflict and violence via country-level, or situational monitoring. That is, we learn of a displacement situation and begin collecting data on it over time.

We have historically published three main figures – the total number of people displaced as of the end of the year, the number of new displacements during the year and the number of people who returned during the year. Where possible, we have also reported on the number of IDPs who have settled elsewhere or integrated locally, those who have sought safety by continuing their flight across an international border and

the number of births and deaths in displacement. We calculate our figures as follows:

New displacement

We may calculate the new displacement inflow for a given year, represented by the orange "internal displacement" arrow in figure A.1, in a number of ways. If our partners provide us with data on new displacement once a year, we simply report the annually aggregated figure. More often, however, they provide us with such data on a monthly or quarterly basis, in which case we publish the sum of the estimates reported.

For Ukraine, we analysed data on IDPs as recorded by the Ministry of Social Policy. These records cover the whole 2016 calendar year, providing the number of people displaced at a given date. Positive differences between two data points give some indication of the minimum number of displacements that occurred in that time interval (see figure A.2).
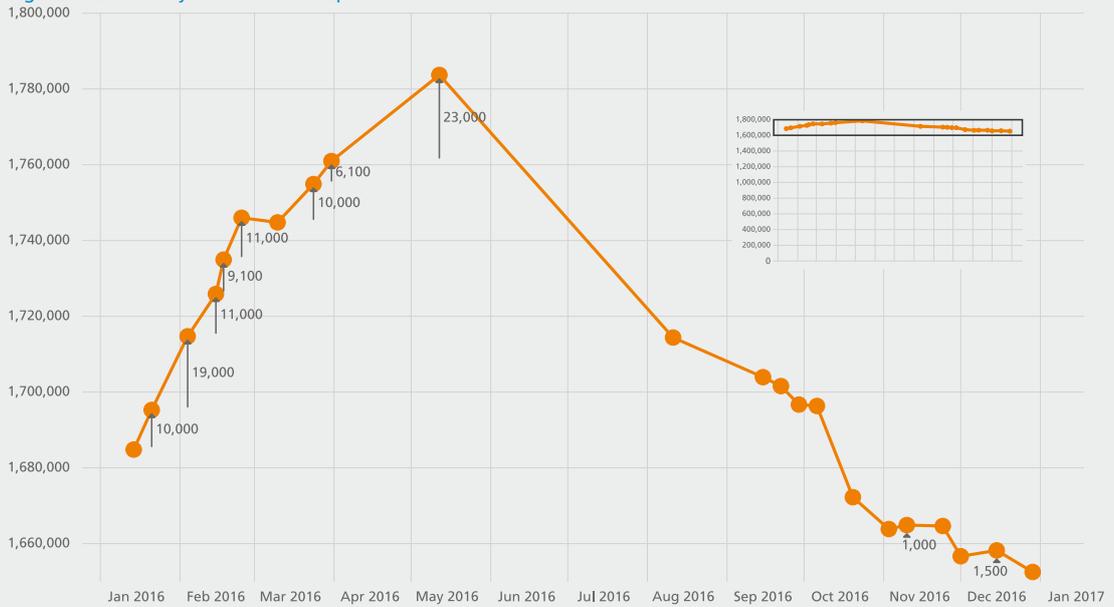
It should be noted that "new displacement" is somewhat misleading in that data may capture the same people being displaced more than once during the year. Given that we are unable to track individual IDPs, it is often not possible to determine the extent to which this is the case for the numbers reported.

The current lack of disaggregated data on IDPs who fail to achieve durable solutions, and on cross-border returns to displacement, also means that such inflows are taken as incidents of new displacement.
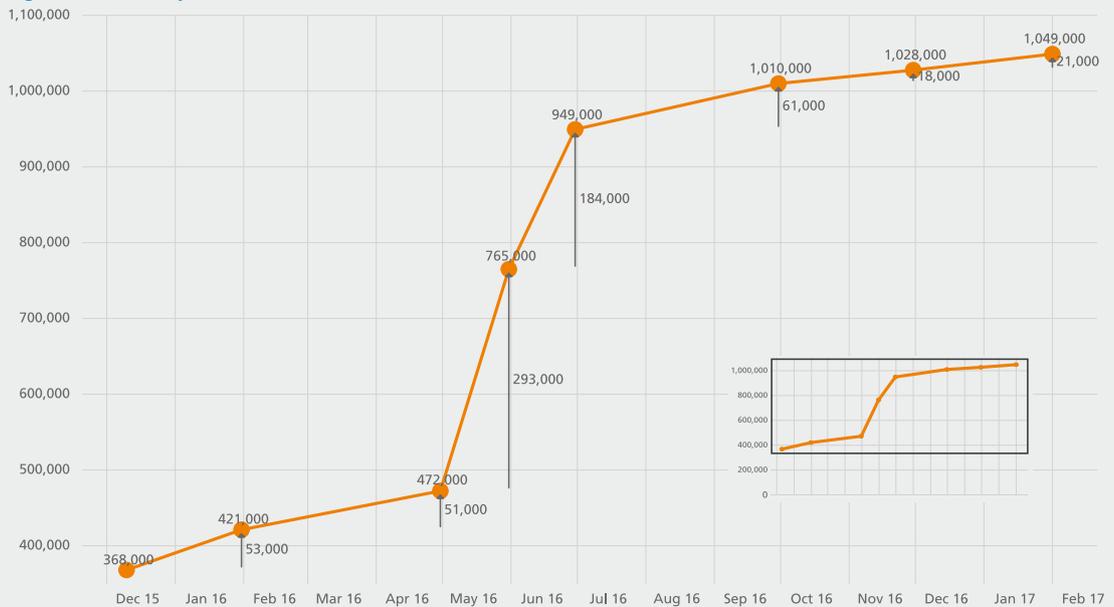
Capturing the end of displacement

We calculate annual return flow estimates in a similar way to those for new displacement. For Yemen, the aggregated return flow for 2016 represents the sum of the reported monthly figures (see figure A.3).

Figure A.2: Monthly data on new displacement in Ukraine

23,000

6,100
10,000

11,000

9,100

11,000

19,000

10,000

Jan 2016   Feb 2016   Mar 2016   Apr 2016   May 2016   Jun 2016   Jul 2016   Aug 2016   Sep 2016   Oct 2016   Nov 2016   Dec 2016   Jan 2017

1,000

1,500

Source: Ministry of Social Policy

Figure A.3: Monthly data on returns in Yemen

1,049,000
21,000

1,028,000
18,000

1,010,000
61,000

949,000

184,000

765,000

293,000

472,000
51,000

421,000
53,000

368,000

Dec 15   Jan 16   Feb 16   Mar 16   Apr 16   May 16   Jun 16   Jul 16   Aug 16   Sep 16   Oct 16   Nov 16   Dec 16   Jan 17   Feb 17

Source: Task Force on Population Movement, TFPM

The same procedure applies to reporting data on local integration and settlement elsewhere, when it is available. It is important to note that accounting for returns, local integration and resettlement reduces the number of IDPs we report, but it does not necessarily mean that they have achieved durable solutions to their displacement. Data to assess the sustainability of these processes is not available at the global level, nor are there universally accepted indicators for measuring their progress.

Cross-border movements

When possible, we deduct the number of IDPs who flee across an international border. In order for us to be able to do this, those collecting information about refugees and asylum seekers need to register whether people had already been displaced prior to fleeing across the border. Failure to do so risks double-counting. The number of refugees and asylum seekers is currently subtracted from their country of origin's general population but not its displaced population.

This year, for the first time, we have accounted for three types of returnees from Pakistan and Iran who found themselves in a situation of internal displacement once (back) in Afghanistan. We included 44,197 Afghans who were deported or voluntarily returned from Iran based on input from our sources in Afghanistan, including UN OCHA. This figure, approximately 10 per cent of the returns from Iran, is predominantly composed of young men who left Afghanistan in search of work and were considered displaced and in need of humanitarian assistance upon their return.

We also included 285,951 individuals, who are part of a significant wave of returns from Pakistan. UNHCR estimates 48 per cent of returning refugees were not able to return to their place of origin. These people therefore fit the government's definition of an internally displaced person.

Finally, we have included a caseload of 22,559 undocumented Afghans who were forcibly deported from Pakistan back to Afghanistan. Given the involuntary nature of the return and the humanitarian needs of these individuals once back in Afghanistan, we consider them to be in a situation of internal displacement.

Births and deaths in displacement

We only account for births and deaths in displacement when our partners provide data. Given the shortage of disaggregated data and the fact that IDPs' fertility and mortality rates may not correspond with national figures, we do not try to extrapolate births and deaths in displacement from national demographic data.

Depending on the scale and duration of displacement, the lack of primary data on these flows can represent a potentially significant blind spot. In protracted crises such as Macedonia's, reported changes in the size of the displaced population may depend more on demographic trends than on returns, local integration and settlement elsewhere, given the lack of progress in these areas.

Total number of IDPs

The inflows and outflows described above all influence the total number or "stock" of IDPs at a given moment in time – 31 December 2016 in the case of this report. We estimate the number of IDPs at the end of the year by triangulating data

reported from one or more sources with a mathematically derived estimate based on the "flow" data available on new displacement, returns, local integration, settlement elsewhere, cross-border flight and births and deaths in displacement.

We arrive at the total number of IDPs as of 31 December 2016 by taking the total at the end of 2015 and adding or subtracting flow data as follows:

$$
\begin{aligned}
\text{Total number of IDPs}_{\text{Dec 2016}} = & \\
\text{Total number of IDPs}_{\text{Dec 2015}} & \\
+ [\text{Births}_{2016} + \text{new displacement}_{2016}] & \\
- [\text{Returns}_{2016} + \text{settlement elsewhere}_{2016} + & \\
\text{local integration}_{2016} + \text{cross-border flight}_{2016} & \\
+ \text{deaths}_{2016}] &
\end{aligned}
$$

The equation is technically incomplete because it does not take into account the "counterflows" represented by failed returns, local integration and settlement elsewhere, or cross-border returns into displacement. Given, however, that data is not collected and these phenomena are accounted for as new rather than repeated displacement, the equation serves its purpose.

In reality, the lack of coverage of the components of our data model and the way outflow data is aggregated mean the actual equation for most countries is often simply:

$$
\begin{aligned}
\text{Total number of IDPs}_{\text{Dec 2016}} = & \\
\text{Total number of IDPs}_{\text{Dec 2015}} & \\
+ \text{New displacement}_{2016} & \\
- \text{Returns}_{2016} &
\end{aligned}
$$

The mathematical formula for estimating the stock of IDPs is at best a modelled approximation. We compare this with the data we obtain from our sources, and they do not always correspond. There are number of reasons for this:

| The initial value – the estimate for the end of the previous year – is incorrect and needs to be revised. This occurs in Afghanistan, among other countries, due to the length of time it takes to verify displacement figures.

| New displacement includes repeated displacement: This is the case every year in countries such as DRC or South Sudan, where pendular displacement – in which IDPs "commute" back and forth between their places of refuge and origin, often to tend to their land – generates higher numbers of displacements that often relate to the same people.

| Double-counting: In Myanmar and other countries in which we compile our national figures from multiple sources, some IDPs may have been counted more than once. We reduce this risk by taking into account the geographical and temporal scope our sources' data.

| Partners change their data collection methodology, as in Ethiopia, or the scope of their geographical coverage, as in Nigeria or Burundi.

| We change our primary source because of the lack of available data or doubts about their credibility, meaning we are working with two very different data sets from one year to the next.

| There is a lack of data on a flow that significantly affects the number of IDPs in a country. Data on the number of refugees and asylum seekers from Syria does not indicate whether they had previously been displaced internally. Similarly, there are indications of displacements in south-eastern Chad as a result of the crisis in the Central African Republic (CAR), but a lack of reliable, updated and verified data.

| Delays in data collection after events leading to displacement toward the end of the year often make it impossible to disaggregate flows by year. In several countries, the year-end figures for 2016 only became available in February or March 2017.

Reflecting the date of sources

When situations remain unchanged from one year to the next, or when flow data is not available, we base our end-of-year estimates on the data provided by our partners. In many countries, however, it has not been updated for several years. In those with complex or multiple displacement crises, such as Chad, Iraq and Myanmar, data for one crisis may be regularly reported, while for others it may be outdated or missing. If there is no credible evidence that IDPs in such situations have returned, integrated locally or settled elsewhere, we have in the past included them in our global figures.

In the interests of transparency, this year's report stratifies the stock of IDPs based on when the primary data was collected (see figure A.4). The length of the bar as a whole represents the total number of IDPs for whom we were able to obtain data. The right-hand section represents data which is increasingly out of date.

## Accounting for displacement associated with disasters

Our estimates for displacement associated with disasters are generated by event rather than by country. We monitor and collect information for all reported disasters from partners including governments, the UN, IFRC and national Red Cross and Red Crescent societies, NGOs and international media outlets. We apply no threshold when doing so, either in terms of the number of people displaced or the distance they have travelled. Our database includes records of one up to 15 million IDPs.

We generate a single "new displacement" estimate for the total number of people displaced by each event. It is important to note that this figure is not necessarily the same as the peak number of IDPs, but instead aims to provide the most comprehensive cumulative figure for those displaced with minimal double-counting.

Figure A.4: Different strata for conflict related stocks of IDPs, ordered by the date of the source data
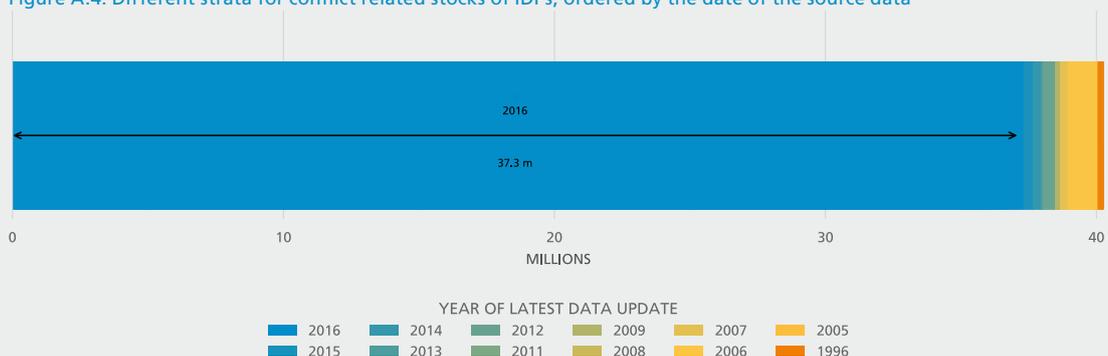
Source: IDMC

We try to collect data from a number of reports on the same disaster, each specifying whether its figures refer to individuals or households, the reporting terms and sources used, the publisher, the title of the source document and the date of publication. When possible we triangulate the figures using competing reports. Sometimes, however, our estimates are derived from a single report. In others, they are the aggregation of a number of reports that together cover the wide geographical area affected.

This dataset allows us to better interpret the context of the figure in each report. In determining our estimates, it is vital that the data selected represents the most comprehensive figure from the most reliable source available for that event at the time when data was collected.

Reporting bias

We are aware that our methodology and data may be subject to different types of reporting bias, some of which are detailed below.

Unequal availability of data: Global reporting tends to emphasise large events in a small number of countries where international agencies, funding partners and media have a substantial presence, or where there is a strong national commitment and capacity to manage disaster risk and collect information.

Under-reporting of small-scale events: These are far more common, but less reported on. Disasters that occur in isolated, insecure or marginalised areas also tend to be under-reported because access and communications are limited.

"Invisible" IDPs: There tends to be significantly more information available on IDPs who take refuge at official or collective sites than on those living with host communities and in other dispersed settings. Given that in many cases the vast majority fall into the second category, figures based on data from collective sites are likely to be substantial underestimates.

Real-time reporting is less reliable, but later assessments may underestimate: Reporting tends to be more frequent but less reliable during 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 considered assessments.

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 did not suffer severe damage but who fled for other reasons.

Our estimates for some disasters are calculated by extrapolating from the number of severely damaged or destroyed homes or the number of families in evacuation centres. In both cases we multiply the housing and family data by the average number of people per household.

Estimating average household size

Primary sources often report the number of homes rendered uninhabitable or the number of families displaced, which we convert into a figure for IDPs by multiplying the numbers by the average household size (AHHS). There is, however, no universal dataset with updated and standardised AHHS data for all countries.

Given the potentially significant influence of AHHS on our estimates, we have continued to update the data and methodology we use to calculate it. This year we used a linear extrapolation obtained with improved methodology developed for the GRID 2016.[3]

The AHHS and therefore our estimates are subject to a margin of error, which means that by applying a particular value we may underestimate or overestimate real figures. If possible we review and update the AHHS every year and, as a general rule, when data is expressed in household or family units, we estimate the number of displaced people according to the AHHS for the year when the data is captured. This applies particularly to figures obtained from historical or retrospective research, notably in protracted or prolonged displacement cases where using a contemporary household size without accounting for demographic changes would have lead to an underestimate for an event that occurred in 2008 (see table A.3).

Table A.3: Illustration of the changes in the AHHS for Benin between 2008 and 2016

For the purpose of the example we use a hypothetical number of 1,000 households displaced	AHHS as of 2008 5.2	AHHS as of 2010 5.1	AHHS as of 2012 5	AHHS as of 2014 4.9	AHHS as of 2016 4.8
Estimated number of people displaced applying the AHHS respective to the year of the figure	5,200	5,100	5,000	4,900	4,800

IDMC's data collection, analytical process, definitions and decision rules

Definition of an IDP

We use the definition of an IDP contained in the 1998 Guiding Principles. The criteria related to the "forced" nature of displacement "within internationally recognized borders" is fundamental in determining whether a person is an IDP, but the Guiding Principles do not set other criteria by which to identity a person fleeing their "home or place of habitual residence".

As such, we interpret IDPs to include not only citizens of a country in which displacement takes place, but also non-nationals such as migrants and asylums seekers in Libya, and Palestinian refugees in Syria and Lebanon; refugees who have returned to their home country but have been unable to go back to their habitual place of residence, such as Afghan refugees returning from Pakistan (see part 2); and stateless people such as the Rohingya.

Forced displacement should not only be associated with the notion of a fixed place of residence, but also flight from traditional "living spaces" that support people's livelihoods, such as pastoralists' grazing areas. Given that the concept of habitual residence is intimately linked to the issue of livelihoods, people who have lost them as a result of their displacement – such as pastoralists in Somalia and elsewhere in eastern Africa – are considered IDPs. We consider a person to be displaced regardless of how far or for how long they flee.

The IASC framework on durable solutions deems displacement to have ended when IDPs have returned home, integrated locally in their place of refuge or settled elsewhere in the country in a sustainable way, and no longer have vulnerabilities linked to their displacement. We acknowledge this concept, but for the purpose of our monitoring and reporting, we do not count IDPs who have returned to their area of origin or place of habitual residence as IDPs, and subtract the figure from our total estimates, whether they are known to have achieved a durable solution or not. This is because in the vast majority of cases it is not possible to properly gauge the extent to which IDPs have achieved a lasting end to their displacement or not.

On the other hand, we consider children born in displacement to be IDPs, and they are included in our estimates. This is particularly pertinent in countries such as Azerbaijan, where displacement has lasted for decades. As such, the number of IDPs in these countries may increase over the years as a result of demographic trends, despite the fact that the original trigger has long ceased to cause any new displacement.

For countries that have been divided into two internationally recognised states, such as Sudan and South Sudan, we do not consider people whose former place of habitual residence is in one of the new entities and refuge in the other as IDPs. For instance, we do not consider a person who fled from what was formerly southern Sudan to northern Sudan an IDP following the creation of South Sudan, but people displaced within either Sudan or South Sudan are considered IDPs.

Data sources

Our ability to report on displacement and provide reliable estimates is contingent on the availability of sources, and their willingness to gather and share data. We draw on information produced or compiled from a wide range of source types. Governments might be expected to have the primary responsibility for counting IDPs, but many others are involved in data gathering, including international organisations, community-based organisations, specialised websites, thematic databases, local authorities, national Red Cross and Red Crescent societies and private sector institutions. Such sources play a significant role,

particularly when governments lack the capacity or will to collect the data or when their estimates are unreliable.

Different sources gather different data for different purposes, with different methodologies and for different objectives. These include operational planning, which is influenced by considerations of timely funding. Divergent objectives often affect the way in which data gatherers estimate target populations or beneficiaries.

We are aware that some sources may also have an interest in manipulating or tweaking the number of IDPs. They may choose to do so in order to call international attention to a crisis, maximise the amount of external assistance received or downplay the scale of a conflict or disaster if the government is held accountable.

In order to mitigate this potential bias, whenever possible we triangulate the data by using several sources and prioritising those we have historically deemed to have been most objective. Particularly for displacement associated with disasters, we monitor the different stages of the humanitarian response cycle, from the emergency to the reconstruction and recovery phase, by identifying the different organisations and indicators that report on displacement over the time.

Language bias also affects our ability to source displacement data comprehensively. We can only obtain and analyse information in the languages we speak and read. Our staff and partners speak most of the required languages, but we inevitably fail to capture some information, particularly for parts of Asia.

Disaggregated data

We seek to obtain not only quantitative data from our sources on possible increases and decreases in the number of IDPs, but also more specific information such as data disaggregated by sex and age (SADD). This is vital in guiding an appropriate and effective response to IDPs' protection and assistance needs.

Relatively little SADD is available for displacement associated with either conflict or disasters. This is mainly because information on IDPs' sex, age and disabilities tends only to be captured in organised settings such as relief camps, while in many cases a significant majority of IDPs live in dispersed settings among host families and communities.

We also aim to gather and report disaggregated information by geographical area and time period in order to paint the most comprehensive and dynamic picture of displacement possible and provide a sound basis for more complex research and analysis.

Even when disaggregated data is available, however, it tends not to represent a statistically significant portion of the overall data collected. More is vital if we are to accurately inform the identification of, and respond to the specific needs of different groups of IDPs.

Methodological challenges particular to displacement associated with conflict

We gather data from primary and secondary sources on the number of people displaced by international and non-international armed conflict and other situations of violence. We aim to include all people forcibly displaced in such contexts.

Sources tend to be numerous during humanitarian crises and visible emergencies, when they compile information to target assistance, as in Syria. During protracted and neglected crises, displacement data tends to be unavailable or out-of-date, as in Armenia, Cyprus, Georgia, Togo and Turkey.

Sources often do not use the same definition of an IDP as the Guiding Principles. Nor do they use the same methodologies, which creates a serious challenge when compiling our estimates. In several countries, including Afghanistan, Bosnia and Herzegovina, DRC, Georgia, Pakistan and Ukraine, only IDPs who have been officially registered with the authorities are counted.

In some countries only one data source is available, while in others there may be several. For each country listed in the 2017 GRID dataset, we systematically looked for several sources. We always strive to identify new data sources, even for countries and situations where others already exist. This enables us to crosscheck, but it may also create confusion because sources rarely explain their methodologies.

When different sources are available, or when a new source provides information, we may still decide to base our estimate on only one

source. That decision may vary from year to year depending on objective criteria, such as their geographical and temporal coverage, or their perceived reliability (see below). We may equally aggregate different data from separate sources to help us extend the geographical coverage of our estimates. As such, our figures are more likely to take into account and reflect both qualitative and quantitative uncertainties.

In many countries affected by conflict and violence, no agencies or mechanisms collect data on the number and kind of people who have sought refuge in urban areas, those who are hosted by relatives or other families or those who have fled to remote areas. This leads to significant underestimates of the number of IDPs.

Data on returns varies significantly from context to context. Sometimes data on returnees is collected after people have returned to their area of origin or place of habitual residence. At other times, our sources use "returns" or "returnees" to indicate that people have departed a location such as a displacement camp with the intention of returning – but with no further information about their location or well-being. In some cases, these returnees may have moved to another camp or become displaced elsewhere, in which case they continue to be counted as displaced. In order to be consistent across all contexts, we subtract returnees from our stock figures. That said, this is a strictly accounting rule and it does not mean that these returnees have reached a durable solution. In order to make that assessment, more follow-up data on returnees is needed.

Selection of countries in the GRID dataset on displacement by conflict and violence

The 2017 GRID dataset contains information on 56 countries and territories. The inclusion of a country is not contingent on a quantitative threshold for the number of IDPs. It depends only on the availability of credible data. The fact that a country is not included does not necessarily imply that no displacement has taken place, but rather that no information has been forthcoming, or that the displacement is not caused by conflict or violence.

Our 2017 GRID estimates include a number of changes from last year's report. They are the result of issues related to the systematic and consistent application of decision rules to all displacement situations, our analysis of the primary causes of displacement, and geopolitical considerations that affect the definition of international borders that are essential to determine whether someone is an IDP, a refugee or stateless. The border issues cover foreign occupation, the creation of new states and unilateral secession.

Geopolitical parameters

We collect and presents data on IDPs for UN members states and other self-governing territories, those with unsettled sovereignty such as the Abyei area, and others with special status such as Palestine and Kosovo. People displaced within areas of an internationally recognised state under foreign occupation are considered IDPs, irrespective of their location with respect to the de facto borders or the territorial claims of the occupying power, providing the original borders still have broad international recognition. Examples are eastern Ukraine, Crimea, South Ossetia and the Turkish Republic of Northern Cyprus.

The inclusion of such countries and other contested territories does not imply any political endorsement or otherwise on IDMC's part.

a. Foreign occupation

We consider people displaced within an internationally recognised state under foreign occupation as IDPs, irrespective of their location with respect to the de facto borders or the territorial claims of the occupying power, providing the original borders still have broad international recognition.

As such, our estimate of the number of IDPs in Cyprus does not only include Greek Cypriots who moved to the southern part of the island at the time of Turkey's invasion in 1974, as was the case in the past. It also incorporates estimates for Turkish Cypriots who moved from southern to northern Cyprus at the time. This interpretation and accounting is consistent with the methodology we have used for other occupied areas, such as Crimea and other parts of eastern Ukraine.

b. Creation of new states

For countries that have been divided into two internationally recognised states, such as Sudan and South Sudan, we consider all people displaced within each of the new entities as IDPs, and we produce separate estimates for each one. People who fled within the previously undivided state and who crossed the border that delineates the new entities are no longer counted as IDPs.

Similarly, we no longer count people who fled from Timor-Leste to West Timor when the former was established in 1999. Their number has been subtracted from our estimate for Indonesia.

c. Unilateral secession

For regional entities such as Abkhazia and South Ossetia, which have unilaterally seceded outside an internationally supported process, we do not count IDPs within them separately from those in the state they have seceded from, in this case Georgia. In cases such as Kosovo, however, where a majority of UN member states have established diplomatic relations with a seceding entity, we do produce estimates for IDPs who have fled within it.

We no longer count people as IDPs if they have crossed what has become a de facto international border and find themselves in different entity from the one in which they were originally displaced. As such, our estimate for Kosovo refers only to people who have fled within the territory itself. Given that the Serbian government reports all IDPs in the country as having come from Kosovo, Serbia is not included in the 2017 GRID.

These decisions not to continue counting people we previously considered IDPs in no way implies that they no longer have vulnerabilities related to their displacement.

Geographical scope and coverage

We aim to capture the full geographical scope of displacement and strive to monitor and report on all situations across the whole of each country we cover. In many, such as Burundi, the DRC, Mozambique, Syria and Turkey, however, data sources do not cover all of the regions where displacement took place. As a result, our figures only reflect geographical areas where humani-

tarian agencies operate, and the objectives of their response and motives for collecting displacement data.

Humanitarian agencies often have difficulty in accessing to conflict zones, which can lead to significant information gaps. Our sources tend to monitor and report on displacement more comprehensively in areas where IDPs are most visible, such as in camps. In most cases, however, agencies fail to record the geographical dynamics of IDPs' movements when registering them. In other cases, such as Myanmar and Syria, they collect data in regions that overlap, often using different methodologies.

Data gatherers are very likely to overlook IDPs living in more dispersed settings. These include people who move to urban areas where they blend in with local inhabitants; those who flee to remote areas, such as the bush in CAR or the forests of Côte d'Ivoire; and those who are hosted by other families or relatives, as in the Philippines. They end up unreported, and the scope and nature of such displacement cannot be quantified and assessed. Their number and fate remain unknown.

Temporal scope and frequency of reporting

The 2017 GRID dataset reports separately on the total number of IDPs as of 31 December 2016, and the number of new displacements during the year. The former reflects the number of people still displaced at the end of the year; the latter includes repeated displacement or other movements of people who fled or returned home during the course of the year.

The figures reported are static, but IDPs' movements are not. For this reason, we aim to improve our methodology and increase not only its geographical, but also its temporal coverage. We plan to produce displacement figures more frequently in order to capture the fluidity and complexity of IDPs' movements.

To do so, we will soon begin piloting a hybrid monitoring methodology that combines event-based and country-based monitoring of displacement situations as they evolve over time (see part 3). The idea is to identify events in near-real time, manually verify those we deem to have led to people fleeing and then engage partners in the field to collect time-series data. In some cases

these partners will help us to identify events that have the potential to trigger displacement by issuing a humanitarian alert.

Methodological challenges particular to displacement by disasters

The 2017 GRID presents our latest findings on new displacement associated with disasters in 2016, and compares it with our historical dataset for 2008 to 2016.

Taxonomic considerations

The 2017 GRID estimates are based on new displacement known to have taken place as a result of disasters for which natural hazards have been identified as the primary trigger. In part 1, we highlighted a number of displacement situations for which it is nearly impossible to identify a single cause or trigger. When available, we use the internationally acknowledged name of hazards and categorise them initially into four main types: geophysical, meteorological, hydrological and climatological. These are then refined into types, sub-types and sub-sub-types (see table A.8).

To better understand the complexities of the phenomena, we plan to break disasters down into various stages and differentiating between their primary, secondary and subsequent triggers.

The 2017 GRID dataset presents figures for displacement associated with sudden-onset hazards, but in future reports we intend to include that associated with slow-onset hazards such as drought. We developed a model-based methodology in 2014, which we used to monitor the displacement of pastoralists in the Horn of Africa during the 2010 to 2011 drought, and we started to collect data on slow-onset hazards in 2015 and continued to do so in 2016.

Temporal coverage

Our dataset records incidents of displacement that occurred in 2016 and are supported by a reliable and comprehensive source. The main challenge we faced in collecting data for the year were overlapping events, such the floods and landslides that occurred in Peru and which we did not include in our estimates because the government provided only an aggregated figure for multiple separate displacements. We have similarly omitted aggregated figures provided by the government of China when we could not trace them back to a specific event.

Table A.8: Taxonomy of natural hazards*

Hazard category	Type	Sub-type	Sub-sub-type
Geophysical	Earthquakes, mass movements, volcanic activity	Ground shaking, tsunamis, sudden subsidence, sinkholes, landslides, rockfalls, ashfalls, lahars, pyroclastic flows, lava flows, toxic gases, glacial lake outburst flows (GLOF), volcanic eruptions	
Meteorological	Storms, extreme temperatures	Extra-tropical storms, tropical storms including hurricanes and cyclones, convective storms, cold waves, heatwaves, severe winter conditions	Derechos, hailstorms, thunderstorms, rainstorms, tornadoes, winter storms, dust storms, storm surges, haze, gales
Hydrological	Flooding, landslides, wave action	Coastal floods, riverine floods, flash floods, ice jam floods, avalanches – snow, debris, mudflows, rockfalls – rogue waves, seiches	
Climatological	Drought, wildfires	Forest fires, land fires –bush, brush and pasture	Fire whirls

* This taxonomy is adapted from the classification system developed by the international disaster database (EM-DAT) maintained by the Centre for Research on the Epidemiology of Disasters (CRED) in Belgium.

Protracted displacement in the aftermath of disasters is also a highly problematic. We produced a first scoping exercise in 2015, which aimed to shed light on the phenomenon by challenging the notion that people who flee a disaster are not likely to remain displaced for long. This false assumption is fostered by only occasional reporting of ongoing cases, often to mark the anniversary of a particular disaster.

Our scoping exercise allowed us to re-examine the issue, and conclude that there are likely to be many more people living in protracted displacement than previously thought. This year, we collected time-series data on the 50 largest displacements in 2016 and the ten largest each year from 2008 to 2015.

Terminology

We use the term "displaced", but it is rarely if ever adopted consistently and unequivocally by different countries or sources (see table A.9). In some countries, such as Afghanistan, the term "returnees" can also refer to IDPs (see part 3). People displaced by floods in 2016 were referred to as "homeless" in DRC and "sheltered" in Saint Vincent and the Grenadines. Often, sources refer to people displaced by disasters as "directly affected". It is true that IDPs are part of a wider population affected by a disaster, but not all those affected are IDPs. As such, additional analysis is required to make sense of the terms sources use, and to understand when and how they signal displacement.

Table A.9: Explanation of reporting terms

Term	Explanation
Displaced	Involuntary or forced movements, evacuation or relocation – when not specified – of individuals or groups of people from their habitual places of residence
Evacuated	Voluntary and forced evacuations, both preventive and in response to the onset of a hazard
Relocated	Voluntary and forced relocations, both preventive and in response to the onset of a hazard
Sheltered/ in relief camp	People accommodated in shelters provided by national authorities or organisations such as NGOs, the UN and IFRC
Homeless	People rendered homeless and without adequate shelter
Uninhabitable/ destroyed housing	Limited to habitual place of residence, and includes houses, retirement homes, prisons, mental healthcare centres and dormitories. The number of destroyed/ uninhabitable houses is multiplied by the AHHS for that country to estimate the number of people who have been rendered homeless and so displaced.
Partially destroyed housing	Data on partially destroyed houses cannot necessarily be taken as a proxy indicator of displacement. This information, however, helps us identify situations we may need to look into further, and access to more detailed shelter assessments are very helpful in this sense. We also use it to triangulate other data. Sometimes, for example, partially destroyed housing is also referred to as uninhabitable.
Forced to flee	To run away from danger. "Flee" implies the forced nature of people's movement and we take it to indicate displacement.
Affected	People whose life has been directly impacted by a disaster or conflict. Displaced people are amongst those affected, but not all affected people are necessarily displaced. There are exceptions, however, and in certain Latin American countries displaced people are referred to as affected for reasons of political sensitivity.
Other	Other indicators of displacement used by local authorities or organisations. They include context-specific terms such as rescued people, people in need of shelter, resettled people and people living in temporary or transitional shelter.

Even within the UN and coordinated international humanitarian reporting mechanisms there is inconsistency in the way different populations are described and counted, with some estimates based on "people affected" and others on "people in need" or "people targeted".

Many terms and expressions are specific to internal displacement, and our database captures the most common ones (see table A.9). They may refer to individuals, groups of people such as families or households, or housing. We use the number of houses destroyed as a proxy because it shows that at least one household has been left homeless. We calculate the number of individuals by applying the AHHS available for each country.

Housing information

Housing information is important in estimating displacement associated with disasters. To produce our 2016 estimates, we analysed more than 300 reports that mentioned housing damage or destruction rather than the number of people displaced. In order to use housing data as a valid proxy, we only consider figures for homes that have been damaged to the extent they are no longer habitable.

Terms that indicate the extent of damage include "houses at risk [of collapse]", "houses severely affected/damaged" and "houses destroyed". We consider housing to be any place where people have established a habitual residence. We include hospitals if the information provided suggests that long-term patients have been displaced.

We also include shelters in refugee and displacement camps. "Collapsed tents" in Jordan's Zaatari refugee camp, for example, are counted as uninhabitable housing. Such cases constitute multiple displacement, in which people have already fled once, only to become displaced again when their camp is flooded.

Evacuation data

We often use data on mandatory evacuations and people staying in official evacuation centres to estimate event-based displacement. This was the case for 8.4 million of the new displacements we reported on in 2016.

On the one hand, the number of people counted in evacuation centres may underestimate the total number of evacuees, as others may take refuge elsewhere. On the other, the number of people ordered to evacuate may overstate the true number, given that some are likely not to heed the order. 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.

Quality assurance and independent peer review

As in previous years, and in order to improve our methodology, we submitted this year's estimates to a quality assurance process to verify the data. The verification stage is as important as the data collection itself, because it allows possible discrepancies to be identified, and the data to be refined before it is finalised. This year's process was mainly led in-house, and all of our entries have been double-checked, through rigorous analysis by experts previously not involved in the data collection and analysis for each of the events.

Colleagues were assigned each country with displacement associated with conflict and disasters involving more than 500 people. They dug through all the data collected and collated by others, asking questions and highlighting potential gaps, and so ensuring the highest possible level of transparency and clarity. As an example of an entry having undergone changes following the internal review process, reports of displacements associated with violence in India were questioned, leading to a rigorous follow-up process with existing and new sources. This allowed us to solidify our data and present it with a much higher level of confidence in its accuracy and value.

Our data on the huge volumes of historical displacement in Colombia also underwent intense scrutiny, including exchanges with OCHA, the government's victims' registry and NGOs. The review unearthed previously unknown information on the primary source's methodology and data treatment processes, which led to significant changes (see spotlight, p.29).

The quality assurance process for displacement associated with conflict was supported by external advice. We presented our figures and methodology to NRC country offices, IOM teams,

UN agencies, government agencies and NGO's in order to benefit from their field knowledge.

In future we aim to extend the disaster verification process to the entire set of annual entries. We have also submitted this methodological annex to external peer reviewers, and elements of our methodology were reviewed in previous years by a different set of independent experts.

We will embed the external peer review and internal quality assurance processes into our future work to ensure that the methods we use to produce our figures are robust and that we have presented them accurately.

Qualitative assessment of confidence in estimates for displacement associated with conflict

Building on lessons from existing assessments

There have been several attempts recently to design confidence assessment schemes to evaluate data on internal displacement as part of a broader movement in the field of humanitarian needs assessments.[5] The Task Force on Population Movement in Yemen (TFPM), for example, has developed a confidence rating based on disaggregation by sex and age, and the availability of data on districts of origin and displacement.[6]

IOM Iraq calculates a confidence rating in order to produce an estimate for each location in its displacement tracking matrix, based on the number of informants used, discrepancies between information from different sources, the accessibility of the location and the ability to independently validate the data received.[7] The Syria dynamic monitoring report (DYNAMO) gives a confidence rating based on the number of sources, the manner and extent that the data can be independently verified, the amount of convergence among the different sources and the degree to which they correspond with contextual information about the situation.[8]

Such assessments may seem reassuring, but if poorly conceived or implemented they may provide a false sense of certainty or confidence. They may hide the arbitrariness of the under-lying criteria and the way they are weighted and aggregated. They may also reflect the biases and challenges inherent in the various steps involved in constructing an index and collecting the data. To limit evaluators' bias and improve objectivity and consistency, clear decision rules are needed that limit the number of dimensions taken into account.

There are ways of overcoming the limitations of points-based scores, but their complexity may render them opaque, adding another layer of potential confusion. Using only four indicators with two to five possible values for each, IOM Iraq's assessment framework yields up to 126 unique possible combinations.[10]

The challenge of applying nationally specific tools at the global level

It is difficult to extrapolate to the global level from confidence ratings designed for national circumstances. The three examples discussed above all refer to situations in which a single organisation or cluster designs the entire national data collection process.

At the global level, aggregation and cross-country comparison is made more difficult by the number of data sources and the fact that their motivations for collecting information ranges from rapid needs assessments to victim compensation without any a priori global coordination. Sources' methodologies also vary widely, from satellite imagery, registration, sampling, key informant interviews and censuses, to name but a few.

This diversity stands in stark contrast to the standardisation of data in the three national examples mentioned above. As such, the same set of criteria cannot easily be used to judge reliability, and the diversity in which the results are reported makes it more difficult to make comparisons between countries.

IDMC's confidence assessment

We designed a comprehensive framework to assess the confidence we have in the estimates we publish. The methodology and results presented in this report are the initial steps of a process we will continue to develop through several more iterations.[4]

Given that we are as yet unable to apply many of the criteria to our data on displacement associated with disasters, we have only assessed our confidence in the figures associated with conflict and violence. In doing so, we applied a common set of criteria based on:

| The methodologies used to collect it
| Whether it could be independently validated
| The degree to which it is geographically comprehensive in terms of the extent of the conflict and associated displacement
| Whether it is disaggregated by sex and age
| The frequency with which it was collected
| How extensively it covers the components of our data model

We have not attempted to weight or rank these factors, nor have we assigned quantitative point values for them or generated an overall score for each source and estimate. In order to do so rigorously, we will first need to empirically test the relative significance of each of the factors.

Some of the data gaps reported can be attributed to the way governments and organisations collect and disseminate data, but this is not always the case. We try to be as comprehensive as possible in our own data collection, but we may overlook some sources that may address the gaps we report. As such, our assessment reflects the level of detail of the data we were able to collect and process from various sources – not the level of detail of all the data that exists or was published by each provider.

Our confidence assessment for the largest stock and new displacement figures associated with conflict is shown below in table A.10. Our assessment for the full list of countries is available on our website.

Notes on IDMC's confidence assessment criteria

Data disaggregated by sex and age (SADD): The availability of SADD does not directly factor into the calculation of the number of IDPs, but it can be considered a proxy for detailed data collection practices.

Geographically disaggregated data: Such data is not, per se, an absolute requirement for accurate national estimates of displacement. In many countries, however, some of the entities that collect data only have access to some regions. Geographical disaggregation allows for triangu-

lation and gaps to be identified, while its absence can lead to possible double-counting. TFPM in Yemen uses a similar rationale in its confidence rating to justify discarding data when location information is incomplete.

Multiple data sources: The availability of data from a number of independent sources does not guarantee higher quality or more accurate overall results. It can, however, prompt discussion of the various estimates available and the methodologies used to derive them. It also sometimes permits triangulation, which is useful in situations for which displacement estimates are highly sensitive or more susceptible to data collectors' biases.

Temporal dimensions: The frequency of updates is a relative criteria. Unfolding crises and rapidly changing situations such as those in Syria, Iraq and Yemen require more frequent updates than stable and often protracted situations such as in Armenia and Cyprus. Yearly updates may suffice for some situations, but for others, it can exclude some of the shorter-term displacements.

Next steps

Our confidence assessment is a work in progress, and we welcome input from partners interested in contributing to its development. For this report we assessed our confidence in all the conflict figures reported. This represents a significant increase with respect to the GRID 2016, where only 11 countries were considered. We plan to apply our criteria to all of the data we receive and analyse so that our estimates are as accurate as possible. In doing so, our data users will be made aware of the magnitude of uncertainty the data contains, and the underlying reasons for it.

Notes

1.   UN, World Population Prospects, 2015 revision, available at goo.gl/TFtEjY
2.   World Bank, Country and Lending Groups, available at goo.gl/vvClzK
3.   IDMC, Global Report on Internal Displacement 2016, May 2016, p.79, available at goo.gl/VOc8OZ
4.   Benini A, Shikh Aiyob M, Chataigner P et al, Confidence in needs assessment data: the use of confidence ratings in the Syria multi-sectoral needs assessment (MSNA), a note for ACAPS and MapAction, April 2015, available at goo.gl/Vo7W01

Table A.10: IDMC confidence assessment of conflict-related displacement figures

New displacements	Democratic Republic of the Congo	Syrian Arab Republic	Iraq	Afghanistan	Nigeria	Yemen	India	Ethiopia	South Sudan	Philippines
Methodology	Registration, key informants, media monitoring	Other	Key informants	Registration, other	Registration, key informants	Key informants, other	Media monitoring	Registration, key informants, other	Registration, key informants	Registration
Data triangulation	Some local triangulation	No triangulation	Some local triangulation	No triangulation	Some local triangulation	Some local triangulation	No triangulation	No triangulation	Some local triangulation	No triangulation
Geographical coverage	All relevant areas covered	All relevant areas covered	All relevant areas covered	All relevant areas covered	Partial coverage	All relevant areas covered	Partial coverage	Partial coverage	Partial coverage	All relevant areas covered
Geographical disaggregation	Subnational - admin 1	Subnational - admin 1	Admin 2 or more	Admin 2 or more	Subnational - admin 1	Subnational - admin 1	Subnational - admin 1	Subnational - admin 1	Subnational - admin 1	Admin 2 or more
Reporting unit	People, households	People, households	People, households	People, households	People, households	People	People, percentage of population	People	People	People
Frequency of reporting	More than once a month	Other	More than once a month	More than once a month	Other	Every month	Unknown	Other	More than once a month	Every month
Disaggregation - sex	Yes	No	No	Yes	Yes	No	No	Partial	Yes	No
Disaggregation - age	Yes	No	No	Partial	Yes	No	No	Partial	Yes	No
Data on returns	Yes	Partial	Yes	No	Yes	Yes	Yes	Partial	Yes	Yes
Data on deaths	No	No	No	No	No	No	No	No	No	No
Data on births	No	No	No	No	No	No	No	No	No	No
Data on cross-border movements	Yes	Partial	No	No	Yes	No	Yes	No	Yes	No
Data on local integration	No	No	No	No	No	No	No	No	No	No
Data on settlements elsewhere	No	No	No	No	No	No	Yes	No	No	No

Stock	Colombia	Syrian Arab Republic	Sudan	Iraq	Democratic Republic of the Congo	Yemen	Nigeria	South Sudan	Ukraine	Afghanistan
Methodology	Registration	Satellite imagery, other	Registration, key informants	Key informants	Registration, key informants, media monitoring	Key informants, other	Registration, key informants	Registration, key informants	Registration, unknown	Registration, key informants, media monitoring
Data triangulation	No triangulation	Some local triangulation	Some local triangulation	Some local triangulation	Some local triangulation	Some local triangulation	Some local triangulation	Some local triangulation	No triangulation	Contradictory data
Geographical coverage	All relevant areas covered	All relevant areas covered	Partial coverage	All relevant areas covered	All relevant areas covered	All relevant areas covered	Partial coverage	Partial coverage	Partial coverage	All relevant areas covered
Geographical disaggregation	Admin 2 or more	Subnational - admin 1	Subnational - admin 1	Admin 2 or more	Subnational - admin 1	Subnational - admin 1	Subnational - admin 1	Subnational - admin 1	Admin 2 or more	Admin 2 or more
Reporting unit	People	People	People	People, households	People, households	People	People, households	People	People	People, households
Frequency of reporting	More than once a month	Once a year	Other	More than once a month	More than once a month	Every month	Other	More than once a month	More than once a month	Other
Disaggregation - sex	Yes	No	No	No	Yes	No	Yes	Yes	No	No
Disaggregation - age	Yes	No	No	No	Yes	No	Yes	Yes	No	No
Data on returns	No	No	Partial	Yes	Yes	Yes	Yes	Yes	No	Yes
Data on deaths	Yes	No	No	No	No	No	No	No	No	No
Data on births	No	No	No	No	No	No	No	No	No	No
Data on cross-border movements	No	No	Partial	No	Yes	No	Yes	Yes	No	No
Data on local integration	No	No	No	No	No	No	No	No	No	No
Data on settlements elsewhere	No	No	No	No	No	No	No	No	No	No