

Where do the crisis indicator numbers come from?

I've been tracking the provenance of some GIS datasets lately, working on the licenses and attributions that have to be attached to them when they're released as open data. I've also been working on automatically generating development indicator sets for a country - the numbers that help responders understand the state of a country *before* a crisis happens (social problems don't go away just because an earthquake happens).

A lot of crismapping comes down to persistence, capability and trust. Trust is a biggie: as someone (Gisli Olafssen?) said, a disaster is not the time to be handing out business cards. We need to trust the people we're working with, the systems we're using (within sensible limits and the occasional online equivalent of kicking the case in the right place), and we need to (again within sensible limits) trust the data we're using. Trust in the data that happens during a crisis has been talked about a lot recently (see conversations about verification, spoofing etc). It hasn't been discussed so much when we're talking about crisis indicators.

First, it's not enough just to have a number, even if it's only used as a rough rule-of-thumb for how bad a prior situation was. We also need to know how much we can trust that number, which means knowing where the number came from: who collected the data, how big a survey it was, how accurate the numbers are likely to be.

So, in recent work, we've started by looking for sources. Most databanks have a list of sources somewhere in the dataset description (or for individual datapoints, in the dataset footnotes); most maps have these somewhere in the margins. If they don't, they should, because they can tell us a lot about how much we can trust the numbers.

Take, for example the MMR, the maternal mortality rate (annual number of deaths per 100,000 live births) for the Democratic Republic of Congo in the last few years. There are several versions of this number, and here's a quick search for them and their listed sources:

Source	Value	Year	Quoted source(s) for the number
CIA World Factbook's DR Congo page	540	2010	No source quoted. But this is the same number as the World Bank indicator SH.STA.MMRT, so it's probably from there.
World Bank Indicator SH.STA.MMRT	540	2010	WDI and GDF 2010. The data are estimated with a regression model using information on

Trends in Maternal Mortality: 1990-2010	540	2010	fertility, birth attendants, and HIV prevalence. Trends in Maternal Mortality: 1990-2010. Estimates Developed by WHO, UNICEF, UNFPA and the World Bank. • Explains in great detail how the number was calculated, and which datasources were used for it. Gives an uncertainty range for the number (300-1100), number of deaths (15000), lifetime risk (30) and PM, the percentage of maternal deaths in deaths of women of reproductive age (18.4%). Also explains that the MMR has been rounded to the nearest 10.
Data.un.org: UNICEF State of the World's Children 2010 report	670	2008	UN_WHO, UNICEF, UNFPA and World Bank . This is probably the same source used in the World Bank figures, but there's no reference to follow on the data webpage. Periodically, the United Nations Inter-agency Group (WHO, UNICEF, UNFPA and the World Bank) produces internationally comparable sets of maternal mortality data that account for the well-documented problems of under-reporting and

Data.un.org: Millenium
Development Goals 670

2008

misclassification of maternal deaths, including also estimates for countries with no data. Please note that owing to an evolving methodology, these values are not comparable with previously reported maternal mortality ratio “adjusted” values. Data.un.org: UNICEF State of the World’s Children 2010 report: MMR Reported 550 2006-2010 UN_Nationally representative sources, including household surveys and vital registration. The maternal mortality data in the column headed “reported” refer to data reported by national authorities. Trends in Maternal Mortality: 1990-2008. WHO/UNICEF/UNFPA/WB. UNICEF State of the World’s Children 2012 report (on unicef.org): Reported rate 550 2006-2010 The maternal mortality data in the column headed “reported” refer to data reported by national authorities. UNICEF State of the World’s Children report 2012 (on

Data.un.org Gender 990 2000
Info: MMR (estimate),
female 15-49 yr

unicef.org): Adjusted rate 670 2006-2010 The data in the column headed "adjusted" refer to the 2008 United Nations inter-agency maternal mortality estimates that were released in late 2010. Periodically, the United Nations Inter-agency Group (WHO, UNICEF, UNFPA and the World Bank) produces internationally comparable sets of maternal mortality data that account for the well-documented problems of under-reporting and misclassification of maternal deaths, including also estimates for countries with no data. Please note that owing to an evolving methodology, these values are not comparable with previously reported maternal mortality ratio "adjusted" values. • Comparable time series on maternal mortality ratios for the years 1990, 1995, 2000, 2005 and 2008 are available at www.childinfo.org. • WHO_Reproductive Health Indicators Database_Jul2007

OverCognition

Journeys through development data.

<http://overcognition.com>

Data.un.org Gender Info: MMR (low estimate), female 15-49 yr	250	2000	(International estimate) WHO_Reproductive Health Indicators Database_Jul2007
Data.un.org Gender Info: MMR (high estimate), female 15-49 yr	1800	2000	(International estimate) WHO_Reproductive Health Indicators Database_Jul2007
WHO Reproductive Health Indicators Database Database not found.			(International estimate) Might have been superceded by WHO's Global Health Observatory.WHO Global Health Repository global burden of disease death estimates by sex, maternal conditions (GBD code W042) 34.7 2008 There's a set of references in the spreadsheet's Notes page, but no specific reference given for this number.

We learn several things from this.

- The numbers are estimates. It's not always clear that this is so.
- Even on the same datasite, numbers for the same indicator can differ and aren't always updated.
- You can sometimes use the numbers to guess at their source (e.g. the CIA figure) - but you have to be careful about generalizing this.

Maternal mortality turned out to have a very strong single source (the trends in maternal mortality reports). But for many development indicators, the numbers from different sources very rarely match, and are sometimes very wildly different. So if we want a number (or even a couple of numbers) to use, we need to do some detective work. More on this soon.