Annex 2

GAR Disaster loss data universe

In order to have a stronger statistical significance GAR11 has incorporated substantial new data to improve the analysis of extensive risk. All of the 12 databases that were used for analysis of GAR09 have been once more reviewed for quality control and updated to include disaster loss data for 2008 and 2009, and 9 new countries have contributed data for the analysis (Chile, El Salvador, Guatemala, Indonesia, Jordan, Mozambique, Panama Syria and Yemen).

The data universe, see Table A-1 now includes 195,558 local level disaster reports, covering a 40 year period from 20 countries (including two Indian states) with a total population of more than 850 million people in 2009. Most of the countries have data starting from 1970, but given some of the new databases have a shorter coverage all analysis have been made using the common period for all datasets, 1989-2009.

Country	Reports	Deaths	Injured	Missing	Houses Destroyed	Houses Damaged	Affected	Population 2009	From year
Argentina	16,211	3,377	22,470	810	53,973	141,381	23,271,305	40,164,561	1970
Bolivia	2,655	1,190	1,133	254	6,249	8,200	832,980	10,187,067	1970
Chile	10,892	3,184	6,811	640	101,877	278,087	8,052,236	16,983,720	1970
Colombia	24,554	35,898	26,447	2812	183,106	681,404	22,688,062	45,103,268	1970
Costa Rica	11,076	516	51	62	8,796	50,800	32,405	4,509,290	1970
Ecuador	4,783	3,019	2,535	1228	12,074	58,785	1,293,799	14,032,233	1970
Guatemala	4,285	1,953	2,789	1113	20,941	105,985	3,339,301	14,009,133	1989
Indonesia	7,098	191,101	317,569	17059	1,078,498	1,113,316	17,808,509	231,298,009	1972
Iran	2,460	137,381	71,145	2501	138,072	325,186	2,684,134	73,736,600	1970
Jordan	444	140	2,181	34	83	582	331,022	6,318,200	1982
Mexico	22,054	31,442	2,882,359	9273	432,812	2,781,635	59,882,327	106,116,969	1970
Mozambique	3,907	106,741	1677	1037	899,442	194,810	42,044,552	21,891,905	1979
Nepal	13,512	11,541	12,446	2689	216,627	159,269	4,666,973	28,294,580	1971
Orissa	9,618	34,787	13,370	1205	1,729,236	3217,877	103,053,490	39,906,920	1970
Panama	3,002	339	1292	39	13,534	70,678	345,782	3,304,461	1989
Peru	15,268	40,994	65,675	9136	438,376	398,237	2,218,035	29,330,481	1988
Salvador	3,366	4,541	15,087	535	180,277	202,701	343,817	7,124,374	1970
Sri Lanka	13,326	33,553	21,645	1983	133,416	345,935	26,632,693	20,476,600	1974
Syria	7,326	679	1,312	0	468	1,311	809,681	20,463,800	1980
Tamil Nadu	13,800	5,610	4,819	3105	272,657	991,548	5,753,375	65,597,936	1976
Venezuela	4,449	3,015	379	1059	56,285	158,288	2,932,101	28,143,584	1970
Yemen	1,472	2,797	1,785,659	287	21,697	36,542	27,044	23,580,000	1989
TOTAL	**195,558**	**653,798**	**5,258,851**	**56,861**	**5,998,496**	**11,322,557**	**329,043,623**	**850,573,691**	**1989**

Table A-1 GAR 2011 disaster loss data universe: 20 datasets from 21 countries/states, all updated and reviewed up to 2009.

GAR Thresholds: data and method

The analysis conducted for the GAR 2009 involved as first step a determination of a threshold under which the impact of a hazard within a geographical unit comparable to a municipality would be considered as Extensive. The process of determining these thresholds involved a research process during which several methods were tested, including a heuristically driven approach, a statistical approach involving determination of outliers, a mathematical approach attempting to obtain maximums or inflexion points of the derivative of the accumulative of fatalities and losses and a simpler but effective method of using percentiles. It's worth noting that all approaches produced the same and consistent results with the sample considered.

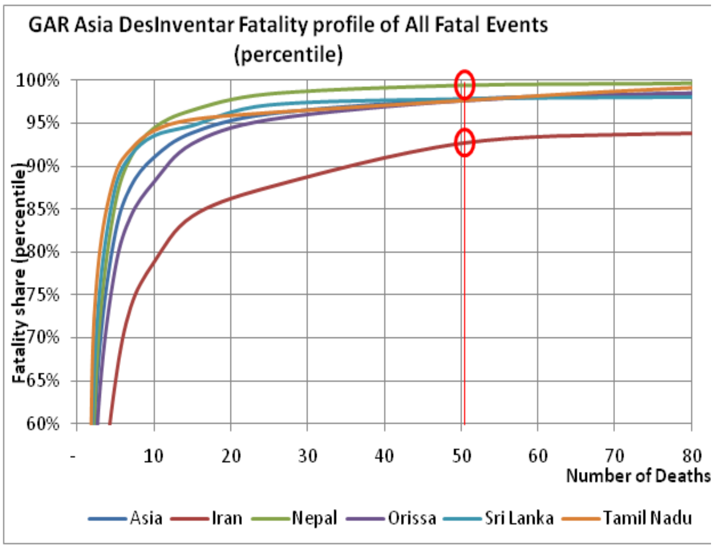

Figure 1. Fatality profile used for GAR 2009 threshold calculation

The latest (and simpler) process used percentiles to set the threshold at the point where the slope of the percentile was below 0.1%. With this threshold (have 50 or less killed), 98% of all reports are left under the Extensive category with only 8.7% of all fatalities, but it can be seen the remaining 2% of the reports concentrate 91.4% of all mortality. The results of this method are shown below in form of a basic table and a chart showing where the cut was made.

GAR 2009 Asia Country Fatality profile (by percentile of All reported & Fatal Events)													
		Asia		**Iran**		**Nepal**		**India -Orissa**		**Sri Lanka**		**India-Tamil Nadu**	
Accum. fatalities	**Class Intervals**	**Fatal events**	**Fatality Slope**	**Fatal events**	**Fatality Slope**	**Fatal events**	**Fatality Slope**	**Fatal events**	**Fatality Slope**	**Fatal events**	**Fatality Slope**	**Fatal events**	**Fatalit y Slope**
	-	0%		0%		0%		0%		0%		0%	
1.4%	1	0%	-	0%	-	0%	-	0%	-	0%	-	0%	-
2.3%	2	51%	0.507	30%	0.303	53%	0.526	44%	0.442	59%	0.589	66%	0.664
3.0%	3	69%	0.183	48%	0.181	72%	0.194	64%	0.197	76%	0.169	80%	0.138
4.0%	5	82%	0.066	65%	0.082	86%	0.068	78%	0.071	88%	0.059	89%	0.044
4.6%	7	88%	0.027	74%	0.044	91%	0.027	84%	0.030	91%	0.019	92%	0.015
5.3%	10	91%	0.011	79%	0.017	94%	0.011	88%	0.013	94%	0.008	94%	0.007
6.1%	15	94%	0.006	84%	0.011	97%	0.004	93%	0.009	95%	0.002	95%	0.003
7.1%	25	96%	0.002	88%	0.003	98%	0.002	95%	0.003	97%	0.002	96%	0.001
8.7%	50	98%	0.001	93%	0.002	99%	0.000	98%	0.001	98%	0.000	98%	0.001
9.5%	75	98%	0.000	94%	0.000	100%	0.000	98%	0.000	98%	0.000	99%	0.001
10.2%	100	99%	0.000	94%	0.000	100%	0.000	99%	0.000	98%	0.000	100%	0.000
12.5%	250	99%	0.000	96%	0.000	100%	0.000	99%	0.000	99%	0.000	100%	0.000
14.5%	500	99%	0.000	97%	0.000	100%	-	100%	0.000	99%	0.000	100%	0.000
16.5%	1,000	100%	0.000	97%	0.000	100%	-	100%	-	99%	0.000	100%	0.000
20.5%	2,500	100%	0.000	98%	0.000	100%	-	100%	-	99%	0.000	100%	-
34.5%	5,000	100%	0.000	98%	0.000	100%	-	100%	0.000	100%	0.000	100%	-
48.6%	7,500	100%	0.000	99%	0.000	100%	-	100%	-	100%	0.000	100%	-
55.7%	10,000	100%	-	99%	0.000	100%	-	100%	-	100%	-	100%	-
76.7%	25,000	100%	-	100%	0.000	100%	-	100%	-	100%	0.000	100%	-
100%	45,000	100%	-	100%	0.000	100%	-	100%	-	100%	-	100%	-
	n' value	**7,614**		**773**		**3,131**		**1,936**		**729**		**1,045**	
		No events											
		Fatal events with asymptotic fatality < 0.1% slope											

Table A1-1: Extensive/Intensive risk threshold calculation for GAR 2009 LAC-ASIA 11 country sample.

The analysis conducted for GAR 2011 revisited this definition of the thresholds between *intensive* and *extensive* risks, looking for less arbitrary, more formal methods to identify *intensive* outliers, taking advantage of this larger sample in number of disaster records but also in number and variety of countries.

Two efforts were conducted by separate teams working on their own methodologies. The first team, from ESCAP (Economic and Social Commission for Asia Pacific) worked on a statistical approach that worked under the assumption that disaster losses are governed by power-law distributions. The second team in LAC (CorpoOsso) worked under the assumption that disaster loss data followed a discrete distribution and used a resampling methodology ('Bootstrap'), and with it tested different alternatives for the definition of thresholds. A method that proved to be less sensitive to weight of extreme values (such as the Indonesian Tsunami) was finally selected.

In general terms both teams searched for the thresholds that would separate the reports in two sets, one which would contain a maximum number of losses in a minimum number of reports (configuring the "Intensive risk" set), and a very large set with the great majority of records and containing a small percentage of mortality and housing destruction losses.

The Bootstrap Methodology, used by LAC team, owes its name and its original formulation to Bradley Efron in 1979. This methodology is the most developed line, both from the theoretical and applied points of view of a variety of techniques which constitute a modern, computer-intensive, general purpose approach to statistical inference, falling within the broader class of so called "resampling methods" .

Bootstrapping is essentially the practice of estimating the properties of an estimator (such as variance) by measuring those properties from an actual sample of the real distribution, in this case the observed disaster loss data. An important feature of bootstrapping techniques is that it's applied to random samples of a completely unspecified distribution – which was the assumed case for the distribution of different impacts from disasters. For the study, the sample is the data from the 21 databases which is anyway assumed to be from an independent and identically distributed population.

The estimation of the desired estimator was implemented by constructing a number of resamples of the observed dataset (and of equal size to the observed dataset), each of which is obtained by random sampling with replacement from the original dataset. The LAC's team target was to estimate the probability of low occurrence (which in a normal distribution occurs at 3 standard deviations), where an extreme event is one that has a probability of occurrence of 0.27%, which means that most data (the extensive risk set) will have a probability of occurrence of 99.73%.

This estimator was calculated resampling 10,000 times the original sample. The median of all estimators was then obtained; this median separates the set of calculated estimators into two equal parts and leaves on each side 50% of the data. The value found for this Median establishes the threshold for each country, continent or whole dataset under study.

In addition to establishing an estimator (threshold) for each country, a bootstrap confidence interval was established using the percentile method, which determines the range of values within which the real threshold of each country should fall with a 95% probability.

Thresholds were calculated for individual countries, continent and the entire sample; a set of methods to blend and combine these indicators, but finally the team opted to choose the estimators obtained directly from the entire data universe as a sample.

The team at UNESCAP followed a different path. Using a probabilistic definition of risk, the study presented a methodology that utilized the concept of Value at Risk (VaR) to estimate the extensive risk, presenting estimates of thresholds of impact of disasters in terms of deaths and houses destroyed, which could be used to identify loss reports related to the manifestation of extensive risk.

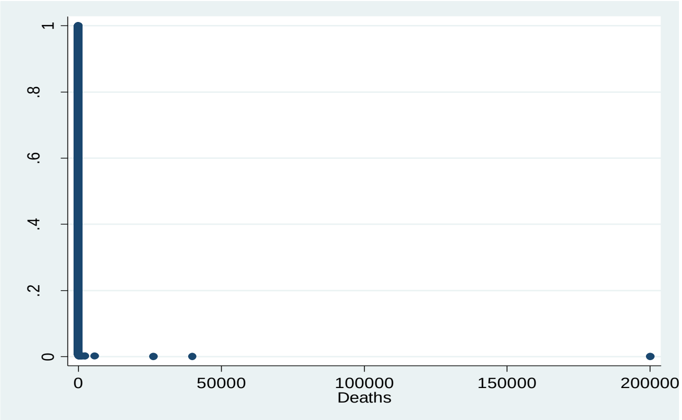

Figure 1. Risk of deaths caused by geological hazards as an exceedance probability, 21 datasets, period 1988-2007

Risk may be expressed in mathematical form as the probability of reaching or surpassing a determined level of economic, social or environmental consequences at a certain place and during a certain period of time.

Figure A-2 illustrates this concept. It shows the risk as an exceedance probability curve of deaths caused by geological hazards in 20 countries/regions for with data is available covering the period of 1988 to 2007.

Given a probability distribution of risk such as the one presented in Figure 1, the risk faced by a country or a region could be presented as the levels of loss that would be exceeded with certain probability during a certain period of time, for example the value of loss per day that has 5% of probability of being exceeded. A well-know measure of risk based on this idea is the Value-at-risk (VAR), which could be defined as the maximum loss during a certain period of time within a confidence interval .

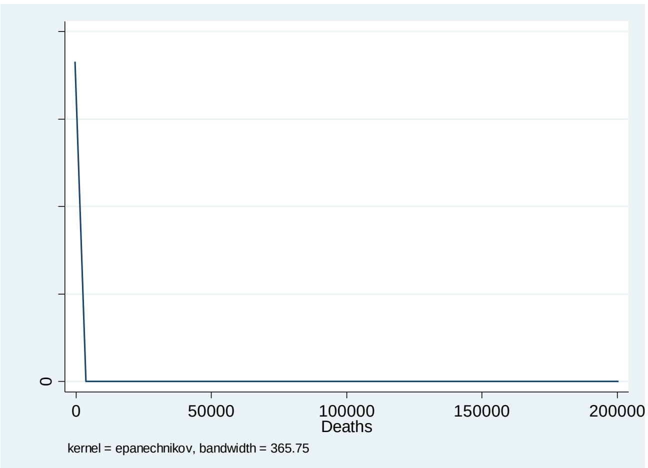

Figure A-2. Probability distribution of deaths caused by geological hazards, Desinventar - period 1988-2007

Once an empirical distribution function of loss is obtained, it should be possible to fit a theoretical probability distribution function and from that estimate the VaR for any confidence interval α in (0,1). The goal in the study was estimating the VaR with a high confidence interval, 95% or more.

Figure 2 illustrates the fact that loss owing to disasters seems to be governed by a fat tail distribution. It shows, for the period from 1988 to 2007, the probability distribution of deaths caused by geological hazards reported in the sample. The number in the horizontal axis at the far right side of the graphic represents the highest loss suffered during that period. When the distribution has the characteristic of a fat tail, the expected size of an event larger than any event yet seen is much larger than the largest event experienced to date (Kousky and Cooke, 2010}

The methodology used to calculate the thresholds for extensive and intensive risk of impact of disasters followed to great extend the methodology adopted by Push (2004) and World Bank, ISDR and CAREC (2009). That methodology has five steps:

1) Assemble the disaster’s impact data by categories;

2) Try to identify the power-law behavior by estimating the scaling parameter (alpha) and the lower bound (xmin);
3) Test the goodness of fit of the power-law (Pareto distribution) when compared with other candidate distributions **for x ≥ xmin;**
4) Estimate the confidence level of the VaR calculated **for 0 ≤ x < xmin;**
5) Find the distribution that better fit the data **for 0 ≤ x < xmin;**
6) Calculate the VaR using the distribution that better fit the data **for 0 ≤ x < xmin; and**
7) Estimate the upper and lower confidence limits of the estimated VaR with 90% confidence.
8) Calculate the thresholds as those values ….{CLOVIS: please complete }

A number of improvements were introduced to the above mentioned original methodologies of WB, ISDR and CAREC, including: testing if a power-law distribution fits the empirical distribution; using maximum likelihood methods for fitting the theoretical probability function instead of regression; using more sophisticated goodness-of- fit tests to select the best-fit distribution; and estimating the confidence interval of the measure of risk and thus of the thresholds.

The thresholds for Intensive risk finally obtained by the teams are as follows:

Mortality: 30 Killed or more in a record

Housing destruction: 600 houses destroyed or more.

Reporting biases on disaster loss data for the period 1989-2009

As has been noted by several authors {REF} potential data biases such as improved reporting were identified but not quantified in GAR09. This field remains still a challenge due to the inherent difficulties in comparing what the real data could be and what proportion of this ideal set is actually being reported.

The challenge has to be addressed on a per country basis as the sources, methodology of historical research and information validation varies on each case.

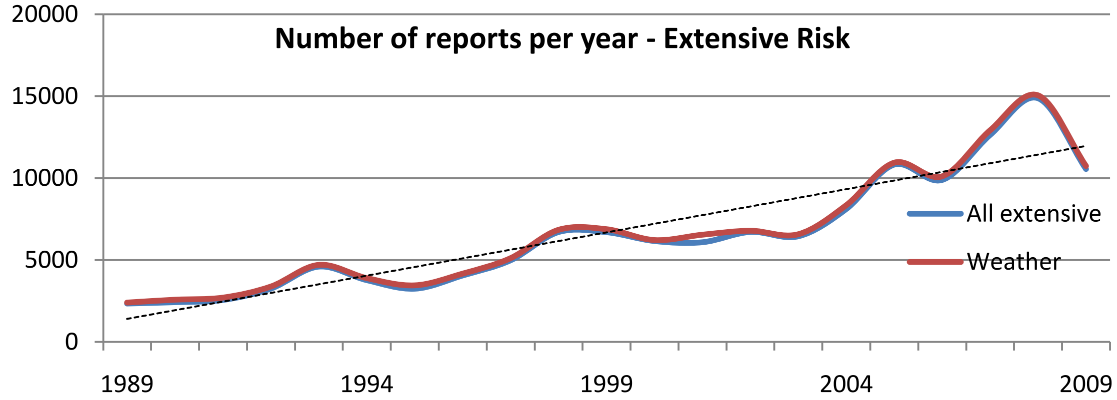

Figure A-3 Trend in absolute number of reports per year in all datasets.

As figure A-3 shows, reporting has grown in the past 20 years from 2,500 reports per years to almost 15,000 in 2008. Figure A-2 shows the same information but per capita.

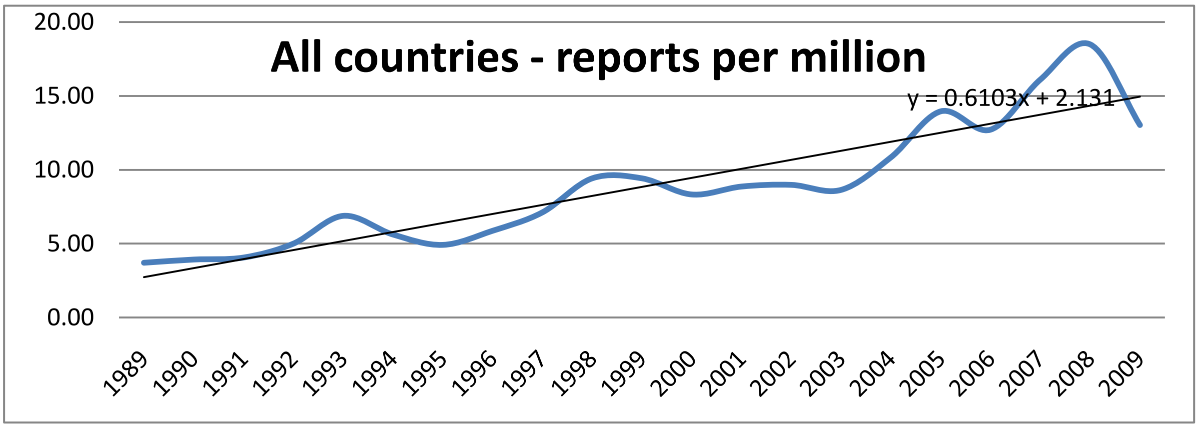

Figure A-4 Trend in relative (per capita) number of reports per year in all datasets.

This, again, suggests there is a large increase in number of reports per capita which can be explained by several reasons which include a higher vulnerability and as suggested, better reporting.

Country	Rate Slope
Chile	-1.06
Peru	-0.21

Jordan	-0.08
Iran	0.00
Guatemala	0.16
Orissa	0.16
Indonesia	0.28
Venezuela	0.43
Argentina	0.53
Mexico	0.60
Colombia	0.62
Tamil Nadu	0.63
Nepal	0.93
Mozambique	0.93
Bolivia	0.99
Ecuador	1.16
Panama	1.19
Syria	2.09
Salvador	2.24
Sri Lanka	4.97
Costa Rica	11.41

However, when things are looked at individual countries, results are surprising. Table A-3 shows the rate at which the number of reports per capita is growing on each country, which would graphically be the slope of the regression line. The overall rate for all countries would be 0.61 more reports per capita each year. This rate has brought the number of reports per million from 3.7 per million in 1989 to the maximum of 18.48 reached in 2008. It's worth noting the steep decrease from 2008 to 2009 due to a 'calm year' with much less occurrence of disasters in 2009 – no reporting bias is believe to happen between these two very recent periods.

Figures A-5 to A6 shows that only for two critical cases (Sri Lanka and especially Costa Rica) a reporting bias is the most obvious and probably only explanation to the huge increase in reporting. In the case of Sri Lanka it can be seen that the increase in reporting starts in 2005, year in which the Disaster Management Center starts activities, triggered by the Indian Ocean Tsunami and much more attention to disasters is paid in the country, resulting in much higher rates of reporting.

The case of Costa Rica is similar as the database has two well defined periods, up until 1997, during which the database was based on media sources, and 1998 and onwards when the CNE (National Emergency Commission) becomes a source to the database and the number of reports per year jumps from an average 150 per year to 700 in the period 1998-2009, with a peak of 1500 reports in year 2005 when the country is hit by three major storms among them Hurricane Stan.

A second group of countries with a high rate of increase in reporting (Ecuador, Panama, Syria and Salvador) present most of the increase in the past 8 years. At this stage the research cannot positively state to what extent this increase (countries almost doubled in number of reports) is due to improved reporting, but intuitively it should have an important influence.

The third group of countries show a low-medium increase of reporting (Mexico, Colombia,Tamil Nadu, Nepal, Mozambique, Bolivia). These countries are in the high risk class of the GAR 2009 and have been effectively subjected to an increasing number of disasters and local studies show that vulnerabilities may be on the rise{REF!}. While it is possible there is some influence of better reporting, the increase should likely be associated with other more important factors. This group's datasets represent about 30% of the total population covered in the sample.

The last two groups, which make for more of half of the countries in the sample and 60% of the population represented, have either a negative growth or a stable number of reports per capita. It is hardly possible then to assign any bias to this sample due to better reporting.

Between the last two groups for which no reporting bias (60%) and very arguably some bias, there is more than 90% of population represented and 72% of all reports in the sample.

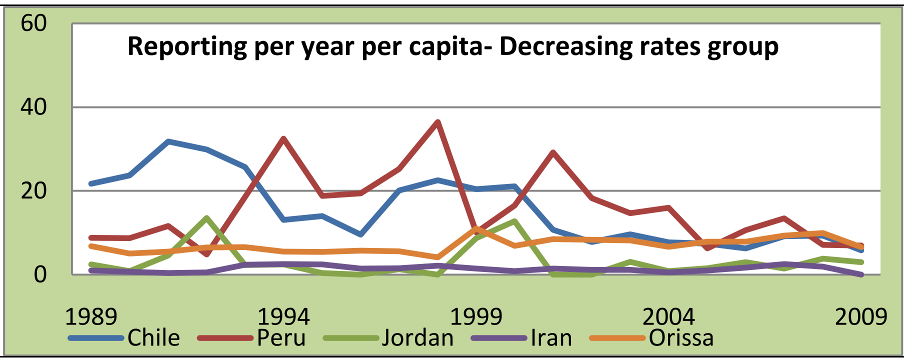

Figure A-4 Trend in relative (per capita) number of reports per year in countries with decreasing reporting rates

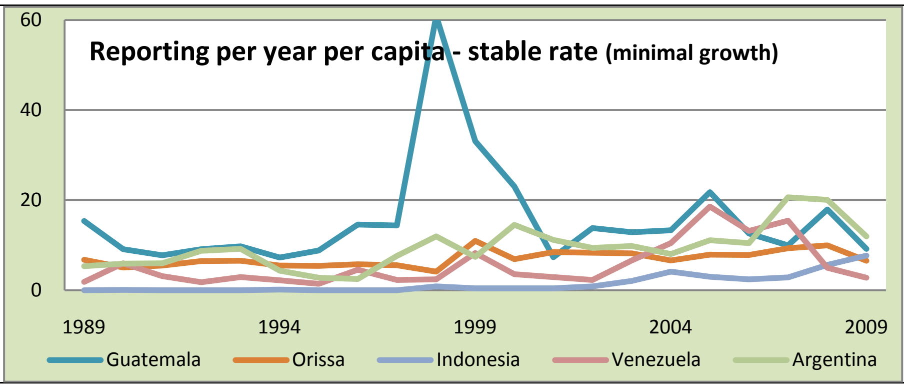

Figure A-5 Trend in relative (per capita) number of reports per year in countries with stable reporting rates

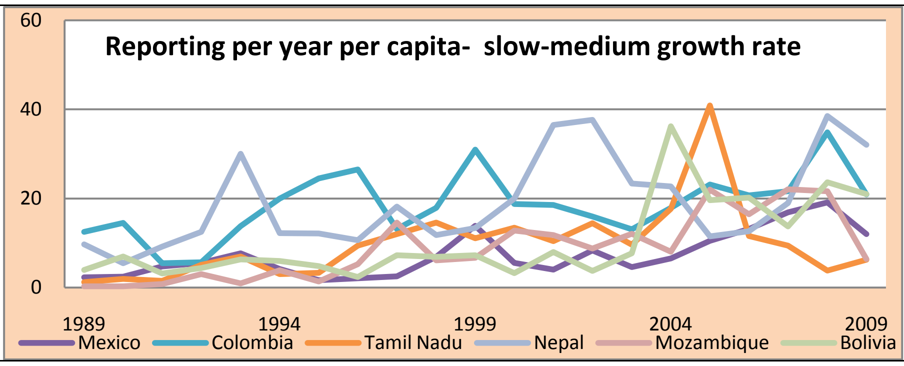

Figure A-6 Trend in relative (per capita) number of reports per year in countries with slow-medium increase in reporting rates

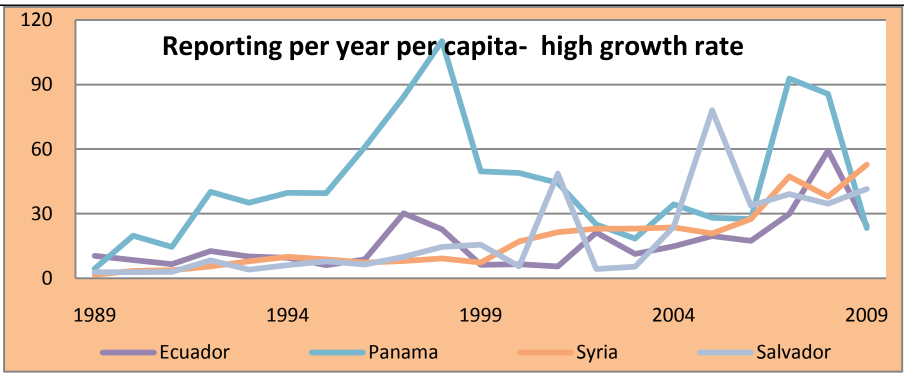

Figure A-7 Trend in relative (per capita) number of reports per year in

countries with high increase in reporting rates. Scale goes to 120 – twice as previous groups.

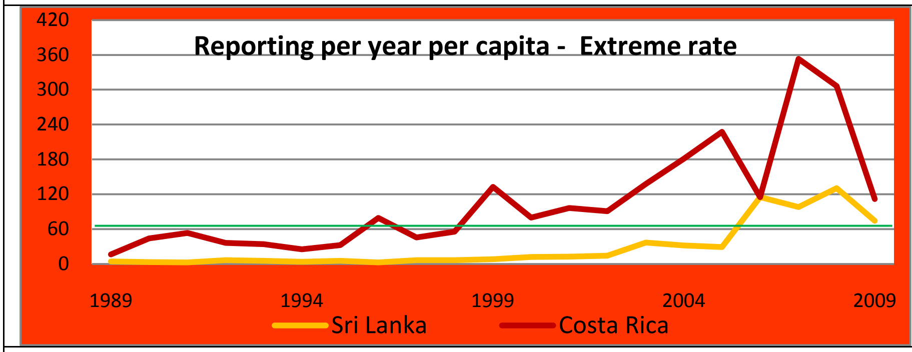

Figure A-8 Trend in relative (per capita) number of reports per year in countries with extremely high increase in reporting rates. Notice the scale in the vertical axis is 6 times the stable group.