Data revolution: Is Latin America prepared and ready to engage?[1]

Suzana Cavenaghi
suzana_cavenaghi@uol.com.br
National School of Statistical Science / Brazilian Institute of Geography and Statistics

In light of the ongoing discussion of the post 2015 Agenda, that suppose to jointly consider the agreements of many recent reviewed agendas as the Environment, Population and Development, Human Rights, among others, data to monitor progress for a better world becomes the star of the time. For developing countries, where economic inequality, internal heterogeneity and cultural diversity are always strong characteristics, the national capacity to produce the needed data for monitoring progress is a challenge that will take more than a large amount of technological advance. The objective of this paper is twofold. From one side will discuss if a handful of indicators of the big basket of the Sustainable Development Goals, proposed by the United Nations Groups set up to discuss the post 2015 agenda, are prone to measure and monitor the set of goals established in the agenda. Mainly, discussing the property of these indicators taking into account an approach of SMART (specific, measurable, appropriate, replicable, and timely) indicators. The second purpose is to analyze if available data from administrative records, surveys and censuses produced in Latin American countries are suitable to produce these indicators. The periodicity and the level of social and geographical disaggregation, utilizing different countries as case studies, will also be discussed. The paper ends pointing out to measures that are needed in order to produce indicators that inform public policies but that are not cumbersome to data producers, taking note of financial and human resources and national capacity available in the region.

Introduction

In the last two years there have been many events related to the revision and evaluation of the latest cycle of implementation of the United Nations (UN) social and environmental conferences. The review of the Millennium Development Goals has, without doubt, been of greatest interest to governments, organizations and civil society because it involves the themes of several agendas (Population and Development, Status of Women, Environment and Development, Human Rights, etc.), in a more or less comprehensive way. In the discussions of the way forward, the lessons learned from the previous 15 years of implementation of the Millennium Development Goals (MDGs) informed the aspects taken into account in the new Sustainable Development Goals. In this sense, since a planned

[1] Presented at the Annual Meeting of the Population Association of America (PAA), San Diego, California, 2015. Special Session: Demography, Demographers and the Data Revolution, Sponsored by International Union for the Scientific Study of Population.

follow-up and monitoring of the MDGs indicators has shown that these are crucial parts of the process, the United Nations Secretary-General set up an Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG)[2] with the objective of recommending how a data revolution should be included in sustainable development monitoring.

In fact, there is a wealth of recent initiatives, recommendations and research on the data revolution[3], which show different aspects of it, although some still clearly confuse a data revolution with the generation and use of Big Data. Specifically, there are initiatives to promote the use of Big Data to produce official statistics, and the United Nations Economic Commission for Europe (UNECE)[4] keeps a registry of the projects carried out by official statistical organizations on projects that involve the use of Big Data in generating official statistics[5].

The final report presented by the IEAG defined the data revolution in a broader sense as:

> *"An explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies such as mobile phones and the 'internet of things', and from other sources, such as qualitative data, citizen-generated data and perceptions data*" (A WORLD THAT COUNTS, 2014, p. 6).

This definition can suggest that the data revolution is a new phenomenon, brought about by a data exchange arising from the new technologies. However, this thinking could be mistaken. Maybe this is a second (or third) data revolution, or even one more wave of the first. Now, for sure, playing a key role is the appearance of Big Data, the huge, mostly

[2] The working process and reports of this group is available at http://www.undatarevolution.org/.
[3] The UN data revolution working group created and has been updating a catalogue with descriptions of and access to these initiatives: http://www.undatarevolution.org/catalog/2/.
[4] The work developed by UNECE is all available at < http://www1.unece.org/stat/platform/display/bigdata/Big+Data+in+Official+Statistics >, accessed March 23, 2015.
[5] Big Data can be understood as structured and not structured information generated in large quantities through the use of different technologies and methodologies needed to handle those quantities of data. According to UNECE, it can be generated in a variety of ways and can be classified, for example, according to the sources of information: Social Networks (blogs, social media, personal documents, e-mails, etc.) generated by humans, Traditional Business Systems (medical records, baking records, E-commerce, Credit Cards, etc.), generated by processes, Internet (data generated by the machines: computers, cell phones, satellites, GPS, etc.) and machine-generated. (UNECE, 2015).

unstructured datasets originating from different sources, especially from the "internet of things"[6]. As stated by the META Group (2001), these are data that comes in volume, velocity, and variety.

Certainly, the data revolution is not a phenomenon of the current decade. According to Friedman (2001), the terminology came about after 1977 with the escalation of large and complex datasets and the development of methodologies for managing a huge amount of structured data produced with the new technologies (the use of computers to gather data versus the old pen and paper questionnaires and forms). However, according to the author, the interest in the use of big data really increased with the appearance of the Data Base Management Systems (DBMS), and started and developed outside, and aside from, the participation of statisticians in the process.

What can be considered new today is the association between the data revolution and the sustainable development agenda. A simple search for the words "data revolution" using the Google search engines gives 387,000[7] results. If the words "sustainable development" are added to the search, the results decrease to 48,000 links, showing that a data revolution is attached to other subjects and may be around for some time. It is interesting to note that compared to other searches, the use of this term is relatively scarce (Or, as some users might say, "data revolution" has not gone viral yet).

Targeting sustainable development within the context of a data revolution is not free of challenges. The use of big data from the Internet of things entails a vast list of ethical, political and economical issues that have yet to be legislated, and which will not be discussed here. Of interest to this paper are the aspects of the data revolution needed to acquire better-structured small data and the big (and structured) data that could come from better data production and the linkage of existing surveys, censuses, and administrative records. The IEAG groups identified important challenges to confront in the data revolution process, including that there are "*not enough high-quality data*" and "*data that are not used*

[6] The sources of the so called 'internet of things' can be generated in several ways. According to the United Nations Economic Commission for Europe (UNECE), on behalf of the international statistical community (2014), it can be divided in three categories: 1. Human-sourced information (Social Networks); 2. Process-mediated data (Traditional Business Systems and Websites); and 3. Machine-generated data (Automated Systems).

[7] A search updated April 4, 2015.

or not usable". In this paper it is assumed that a first step to take in this new wave of the data revolution is to focus attention on these aspects mentioned by the IEAG, primarily, that data producers should take full advantage of new technologies, first, to produce better data with the existing sources of structured data, which would allow proper follow-up of policies, programs and actions. Besides, everyone knows that most big data cannot be cleaned up without the existence of "real data".

In addition to the challenges noted by the IEAG is the fight against poverty. This is one of the main goals of public policies, and often involves societies with great social inequalities. Consequently it is an endeavor that can mean a lack of financial and human resources. The design and implementation of policies in these settings mean that the responsibility, efficiency, efficacy and effectiveness of the programs are at the core of public policies, to ensure progress and government accountability. Hence, the opportunities inherent in this new wave of the data revolution are welcomed and needed, and the involvement of all actors, mainly statisticians, demographers and other data analysts, is very important. This is not only due to the need for the development of methodologies to deal with complex unstructured data sets, which come from the internet of things, but also because several parts of the world are still lacking or have poor quality basic data. For example, it is difficult to know with certainty how many people are born, how many died and from what causes, and still more basic, what is the true age and sex structure of the population and the age structure of deaths. Obviously, these problems involve the poorest regions of the world, and where live people with greater demands. As the IEAG rightly points out:

> [...] **too many countries still have poor data, data arrives too late and too many issues are still barely covered by existing data**. For example, in several countries data on employment are notoriously unreliable, data on age and disability are routinely not collected and a great deal of data is difficult to access to citizens or is not available until several years have passed since the time of collection (A WORLD THAT COUNTS, 2014, p.11).

Beyond the production, dissemination and use of data, a well-known problem with monitoring goals and targets, as is required for the sustainable and development goals, is the appropriate definition of indicators suitable for monitoring progress. Most of times, the main problem of an indicator, and what demographers seek to make the best it can be, is the denominator, or the population exposed to the events being analyzed. Often, the

denominators are not available and one of the tasks of a demographer[8] is to find the best proxies for the population exposed to risk. From the other end, the numerator is not always easy to define, but it depends much more on the correct definition of the event and exposure time than a methodological problem. A big challenge in selecting indicators to monitor a goal or a target is when they are not accurately defined, do not specify the exact target (population), and their time periods for achieving the goal and the expected degree of change are lacking or ambiguous.

In light of these circumstances, the objective of this paper is twofold. First is a discussion of whether a list of indicators, selected from the Sustainable Development Goals (which were proposed by the United Nations Groups, and set up to discuss the post 2015 agenda), are well defined for monitoring the set of goals/targets established in the agenda. Mainly, discussing the properties of these indicators, taking into account the SMART approach (specific, measurable, appropriate and timely). The second and main purpose is to discuss whether the available data from administrative records, surveys and censuses produced in Latin American countries are suitable for producing these indicators. Finally, this paper discusses how governments, non-governmental organizations and multilateral institutions should take the advantage in establishing the grounds for a data revolution that permits an adequate monitoring of the world's progress in the framework of sustainable development.

Indicators of sustainable development

In the year 2000, almost a decade after several major conferences, including the United Nations Conference on Environment and Development, International Conference on Population and Development, and World Conferences on the Status of Women, 189 countries signed the Millennium Declaration, which gave rise to the MDGs and its targets. These goals are established in an official list[9] (since its first publication the goals and indicators have changed) including eight goals, 21 targets and 60 indicators to be followed nationally, and with the recommendation that the indicators should be disaggregated by sex

[8] For a statistician, for example, there are no proxies. It should be the population exposed to the risk at the beginning of the period of exposure, in order to obtain the probability of occurrence in that given period.

[9] The official list of targets and indicators is available at <http://mdgs.un.org/unsd/mdg/Resources/Attach/Indicators/OfficialList2008.doc >.

and urban/rural when possible. In the follow-up marking the 15th anniversary of the Millennium Declaration, the MDGs will be replaced by a much more extensive, complex and challenging set of goals and targets, the already widely known Sustainable Development Goals (SDGs).

The UN Open Working Group on SDGs, which was established to propose the goals to be negotiated among the State Members at the UN General Assembly, drafted 17 goals and 169 targets, the total number of actions under the specific goals. Even if just one indicator were proposed for monitoring each target, the total number would be almost triple that of the previous one. Although there are several lists of proposed indicators, which came from different groups, the UN Statistical Commission was called to present the official list, which will also be negotiated at the 68th Session of the UN General Assembly in September 2015. The Commission decided to work with groups of experts in each area, and asked for a maximum of two indicators linked to each target. The official list of the Commission has not yet been finalized, but its first draft, made available in February 2015, includes 310 indicators, and some of them recommend disaggregation by specific demographic groups[10].

It is important to note that many of the MDG targets set in 2000 were not met by several countries by the 2015 deadline. Now the SDGs will replace some of these goals, repeating those previous targets. In other words, it becomes an extension of the goals without considering any "accountability penalties" for the targets not met. One good example is the target for access to universal sexual and reproductive health and rights (which this paper will analyze): "*Achieving universal access to reproductive health by 2015, as set out at the International Conference on Population and Development"*. It is important to note that not all goals had an explicit deadline as defined by the MDGs, and this is also true for the SDGs. Thus, in the discussion of indicators, it is essential to address first this characteristic of the targets, that is, the time necessary to achieve them.

As for the properties of the indicators, it is suggested that these follow the SMART approach as much as possible, since not all features can be found in the same indicator. In public

[10] The proposal of the Open Working Group suggested that the indicators for monitoring the SDGs should be available by "*income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts*" (PROPOSAL OF THE OPEN WORKING, 2014, p.4)

policy implementation and monitoring, this acronym is now widely known, although some of the characteristics might vary from one study to another. In this approach, the S stands for specific, M for measurable, A for appropriate, R for replicable, and T for timely. That is, the indicator should be specific to the event under consideration (and not encompass other events); it should be an event that can be measured (quantitatively or qualitatively); it should be correctly estimated (directly, not from models); it should maintain a large series for comparison (replicable at different times); it should be measured at a proper time (according to the pace of improvement/deterioration).

In order to adopt this approach, the target itself must also be well defined in these terms, at the risk of being unable to define an adequate indicator. The SMART approach to define goals was first proposed by Doran (1981), and was focused on the management field. The original definition had some criteria that differ from what is used currently, or in different fields. The criteria for a good definition are Specific, Measurable, Assignable, Realistic, and Time-Related. Two other criteria, incorporated later in the proposal, and welcomed in the SDGs process, are Evaluated and Reviewed, which is self-explanatory in public policy contexts, and improves the acronym to SMARTER. These features will be discussed in the last section of this paper.

A Look at Specific Indicator Characteristics

Target 3.1

The health-related objectives in the MDGs were separated into three different goals and six targets. Now, the proposal for the SDGs includes health and well-being in one goal, but the number of targets increased to 13, including broader actions (*Goal 3: Ensure healthy lives and promote well-being for all at all ages*). The first, **Target 3.1** of this goal, deals with maternal mortality and says: "*By 2030, reduce the global maternal mortality ratio to less than 70 per 100,000 live births*". It can be seen at first glance that this target fulfills some important criteria needed for a good indicator: it is specific to maternal mortality (all death related to the period before, during and after birth); it is measureable if there is a list of causes of deaths that can be linked to a woman's death; it is replicable if the same method is applied over time; and there is a time set to achieve the proposed target. When it comes to

appropriate or attainable features, there are some problems with information that some countries have regarding health access for all, but the target in general seems to be well defined. And indeed is so specific that the target itself already defines an indicator to be tracked: the maternal mortality ratio.

What seems to be the problem here? First is that the target in itself sets a goal to be achieved globally. If the actions and policies are designed and implemented nationally, the target does not lead to achievable goals by the end of the period. Currently, there are countries with a maternal mortality ratio much lower than 70 per 100,000, and others with ratios above 200 and even higher. On average, the world overall could achieve this goal, but some countries could have extremely high maternal mortality. Or there is something wrong with the definition of this target. Another problem is that maternal mortality ratio does not specifically establish the population at risk, which should be pregnant women, that is, the denominator does not target the population at risk, which has only a proxy, births, which can be closer to or further from the target, depending on the rate of abortion (induced or not).

It is important to note that the goal set 15 years ago regarding maternal mortality (*Reduce by three quarters, between 1990 and 2015, the maternal mortality ratio*) was not met by many countries. Also, most countries in Latin America do not have accurate data on maternal mortality, even with all the methodological approaches suggested by the Interagency Group for the Estimation of Maternal Mortality and implemented in several countries. In Latin America and the Caribbean, for example, a pilot study carried out by the Economic Commission of Latin America and the Caribbean (ECLAC) pointed out several critical problems in eight countries in the region regarding data collection on maternal deaths (Cobos, Miller, Ruiz Salgueiro, 2013). Indeed, in some countries the number for maternal mortality is obtained using models, by estimating the percentage of deaths of women aged 15 to 49 due to maternity-related causes. This was the case of Brazil, which supposedly has a good statistical system.

The second indicator proposed for this target says "*Skilled birth attendance*". This indicator certainly is correlated with maternal death, but the indicator is not specific to the target. There are countries where this indicator reaches values of almost 100%, however, maternal mortality is high because of low quality of the services provided. An example, but not the

only one, again is Brazil. Indeed, this goal would be achievable by 2016, since maternal mortality is below 70 per thousand (though very high by WHO standards) and the skilled birth attendance is among the best, since 98% of births occur in hospitals. This contradiction suggests that this indicator should be more specific to better relate to the goal.

Target 3.2

The subject of child mortality is dealt with in Target 3.2 "*By 2030, end preventable deaths of newborns and children under five years of age*". In terms of the definition, this target is well set, although it includes two different population groups. However, it would not be impossible to define good indicators to track the trends concerning these events, if the intention is to have indicators of results and not processes. Also, to "end preventable death" seems to be an ambitious target, although it would be desirable.

The indicators proposed by the Statistical Commission under Target 3.2 are 1) Under-five mortality per 1,000 live births and 2) Neonatal mortality per 1,000 live births. The target population is preventable deaths, nonetheless, the first indicator is not specific to it, indeed, a large number of deaths are most difficult to prevent (assuming that a list of preventable deaths is agreed upon by specialists). Moreover, another problem is that there is not an explicit level to be achieved (as there is for the infant mortality rate in the MDGs), although the group that proposed the indicator recommended that this indicator should be below 25 per thousand live births in every country.

The second indicator undoubtedly correlates to preventable deaths, because it refers to deaths that occur in the first 28 days of life, and generally those are highly correlated to death that could be prevented with good prenatal care and quality health assistance during and right after birth. Nonetheless, there are a number of deaths in the first days of life that cannot be considered preventable, not even with a good care. Also, with increasing medicalization and improved technology, some are able to survive after the first month, but do not make it to age 5, or even one year. The correct indicator should count death events that are most likely preventable, however, this requires a very good vital statistics registration system, which most low- and medium-income countries currently do not have.

Target 3.7 and 5.6

Additional targets related to a demographic component are Targets 3.7 and 5.6. The first states "*By 2030, ensure universal access to sexual and reproductive health-care services, including for family planning, information and education, and the integration of reproductive health into national strategies and programmes*". The second, Target 5.6 reads "*Ensure universal access to sexual and reproductive health and reproductive rights as agreed in accordance with the Programme of Action of the International Conference on Population and Development and the Beijing Platform of Action and the outcome documents of their review conferences*". The expansion of the scope of these targets, compared to the related MDG is very welcomed, mainly because the MDG target was set only five years after the Millennium Declaration and simply stated: "*Achieve, by 2015, universal access to reproductive health*". The target previously had four indicators, and the SDGs, in addition to recognizing a broader goal, maintained four indicators for monitoring.

The two first indicators, related to Target 3.7, which is about access to health-care services in reproductive planning, concern adolescent fertility rates and the demand for modern contraception. These same indicators were used in the MDG targets, though now the age range has been extended, adding younger adolescents, aged 10 to 14 years old. Both indicators are related to the level of health-care access to contraception services. However, since the target is so broad, indeed it does not hold up as a SMART target and, thus, the indicators proposed for monitoring will not capture the progress implied in the target. Indeed, in the way it is stated, the target should be a general goal with several targets assigned to it. Also, it is anticipated that in some countries it will be very difficult to collect information on childbearing by women aged 10 to 14. This information, again, would need to come from good vital records.

As for the indicator "demand satisfied with modern contraceptives", it was the reverse in the MDGs, the unmet need for family planning. The change in the approach for the indicator is welcome, but more than a change in the approach, what is needed is a change in the definition of the indicator, mainly regarding the numerator and the denominator. Generally, it includes women of reproductive age (15 to 49) who are married or in a union. Nonetheless, it leaves out large numbers of the female population who are sexually active but are not

married or in a union. The differences can be significant among socioeconomic groups, for example, and it can be a source of enormous differences between and even within countries. Also, less of a factor but still important is that not all women who married or in a union are exposed to pregnancy (or can be accounted as sexually active).

Another important aspect that should be considered in defining the satisfied demand for contraceptives, and which calls for better data collection, is the use of modern contraception. Although it has already been agreed what is considered modern contraception, among different segments of the population the consistency of use varies greatly. Generally, surveys collect information of modern contraception use in the last year, without identifying a possible window for contraceptive failure due to misuse or inconsistent use. In this sense, again, Brazil is an iconic case. There is currently a high rate of modern contraception use, thus a very low unmet need, which is about 6%. However, the level of unwanted (18%) and mistimed (28%) births totals 46% of all recent births (and also large portion of induced abortion), according to the most recent available survey, conducted in 2006, which indeed is already outdated. The high level of adolescent fertility is one of the consequences of inconsistent use of contraceptives in Brazil.

Target 5.6, meanwhile encompasses an entire chapter of the Program of Action of the ICPD agenda. It is additionally pleasing to note that this target was included, because it is within the general goal of achieving gender equality. At the same time, it suffers from the same problems as Target 3.7, because it is very hard to select indicators that are suitable for monitoring progress towards the goal, which has several nuances. First, it focuses on two different aspects of access, one on sexual and reproductive health and another on reproductive rights. Second, because it does not set a timeframe to reach the goal and does not have a target population, it implies that it should be for all populations. The indicators selected, then, reflect this incomplete definition of the target, which is more like a goal.

The first indicator proposed for monitoring Target 5.6 is "*Percentage of women and girls who make decisions about their own sexual and reproductive health and reproductive rights by age, location, income, disability and other characteristics relevant to each country*". What are the parameters that the evaluators will look for in this indicator? Is there a general agreement about this? Indeed it should be 100% free decision making, but the document

about the background information for the selection of indicators does not clarify this point. More worrisome is the fact that the definition of the indicators mixes up several different aspects of decision-making and does not define the age range of the female population selected as the target. Also, it uses a Demographic and Health Survey question on decision-making that is broader than sexual and reproductive health (SRH) and reproductive rights[11].

The proposal of the second indicator puts priority on the legal aspects that guarantee informed choices about the subject of the target (SRH and SR). The indicator is stated as follows: "*The existence of laws and regulations that guarantee all women and adolescents informed choices regarding their sexual and reproductive health and reproductive rights regardless of marital status*". The scope of the indicator seems to include male adolescents, although it is not explicit, but at least does not make SRH and SR a female matter only. However, as for the criteria of a good indicator, one that can really monitor progress, it leaves enormous doubts about its suitability. Also, in the instructions for the indicator there is a long list of types of laws and regulations, without any further definition, since it is a new indicator that still needs to be developed, tested, parameterized, etc.

There is an intention to include more indicators in this analysis, to show other problematic aspects in the selection of the proposed indicators. What follows is a discussion of the state of the art for the quality of data needed for a positive outlook in Latin America, renewing the discussion of what kind of data revolution is urgently needed in the region if the SDGs are to be monitored.

Does Latin America currently produce the data to monitor sustainable development properly?

As mentioned in the previous section, it seems clear that not all Latin American countries are prepared to provide data for the proposed indicators, and even less so for a better list of indicators needed to appropriately monitor progress toward the SDGs. And it should be kept in mind that only a few indicators were select, and those are well known and have been in

[11] The definition of the indicator is as follows: "*This indicator measures different dimensions of choice for women/girls either by themselves or jointly with a spouse or partner, including autonomous decisions to go out of the house; to visit friends and family; to go to a health centre; to choose a particular contraceptive method; and to choose other sexual and reproductive health services.*".

use for a long time. There is also a handful of indicators for which even developed countries lack a good system for collecting the necessary data, mainly those related to environmental processes.

Among the main sources of data in Latin America are the demographic censuses. The region has a long tradition of collecting data on population and its main components, fertility, mortality and internal migration, using methodologies developed in the middle of the last century, with indirect methods. These data are essential as the basis for several indicators proposed in the SDGs. The extended experience in data collection using censuses methodologies has provided Latin America with much better data than many other developing countries have. Indeed, the overall estimate of data omission for the region in 2010 was around 3.3% of the population (Table 1).

However, there are still several errors in the data, and the quality varies from country to country, or across years within a country. And, contrary to what should be expected, data quality does not always trend toward improvement. An example is the region's 2010 round of censuses. Table 1 presents the percentage of population omission, based on the comparison of the census results and population projections[12]. As the Table shows, Paraguay and Chile had severe problems with their last census, with a sub enumeration estimated at around 26.0% and 9.3%, respectively. Those figures for census omission, after decades of experience in data collection, are inconceivable, moreover, because the omission probably has large differentials in terms of socioeconomic and regional characteristics.

Indeed, it should be pointed out that even these figures of omission have errors attached to them because there is no systematic methodology for estimating census coverage in the region. The best way to achieve this would be to have a good base of vital statistics, with close to 100% coverage, and good records of international migration. Nonetheless, as will be analyzed later, there are deficiencies in vital statistics registration in most countries in the region. Another method for estimating census coverage should be post enumeration

[12] One should keep in mind that neither data set can be claimed as the best or the most certain. Censuses have natural omissions (or over enumeration) that are intrinsic to a complex data collection process, and projections have estimation errors based on the parameters used, generally, rates of fertility, mortality and migration by age and the baseline population.

surveys. However, the few countries that have conducted these surveys either do not publish results consistently or do not carry out the surveys consistently as part of the census operation.

Table 1: Implicit census omission according population estimates and projections by census period, Revision of 2012, Celade/ECLAC, Latin American Countries

Country	Decade of the census (census years vary by country)						
	1950	1960	1970	1980	1990	2000	2010 [a]
Latin America	6.3	5.3	5.4	4.2	4.3	3.8	3.3
Argentina	7.2	3.5	3	1.2	1.3	3.2	1.1
Bolivia	0.9	-	6.2	-	8.5	3.8	2.3
Brazil	4.0	4.2	3.6	2.7	3.7	2.8	2.4
Chile	6.3	4.5	7	1.8	2.4	4.3	9.3
Colombia	8.9	3.2	10.6	8.2	6.7	4.9	-
Costa Rica	17.0	9.1	4.2	8.0	-	3.1	6.2
Cuba	6.6		2.0	1.8		0.2	0.1
Ecuador	8.5	8.0	3.4	4.3	5.5	6.0	4.5
El Salvador	15.7	11.9	7.5	-	7.3	6.2	-
Guatemala	10.3	6.8	11.9	15.3	14.2	6.1	-
Haiti	4.1	-	10.1	15.7	-	6.0	-
Honduras	8.1	8.6	11.2	8.3	-	4.9	-
Mexico	8.9	9.8	8.0	5.0	4.3	3.9	3.0
Nicaragua	18.4	20.8	23.8	-	6.3	5.7	-
Panama	13.3	12.5	6.3	8.1	6.2	6.9	7.3
Paraguay	10.8	10.4	9.8	11	7.6	7.8	26.0
Peru	-	3.3	3.0	4.5	3.0	-	3.3
Dominican Republic	10.3	7.8	9.7	7.3	4.7	3.8	5.2
Uruguay	-	1.8	1.5	2.1	2.5	2.1	3.9
Venezuela	3.0	3.6	4.6	7.5	9.1	7.8	6.5

Source: Authorized reprint. Naciones Unidas, Celade/Cepal: Santiago de Chile, November 2013. Principales cambios en las boletas de los censos latinoamericanos de las décadas de 1990, 2000 y 2010, Serie Manuales.

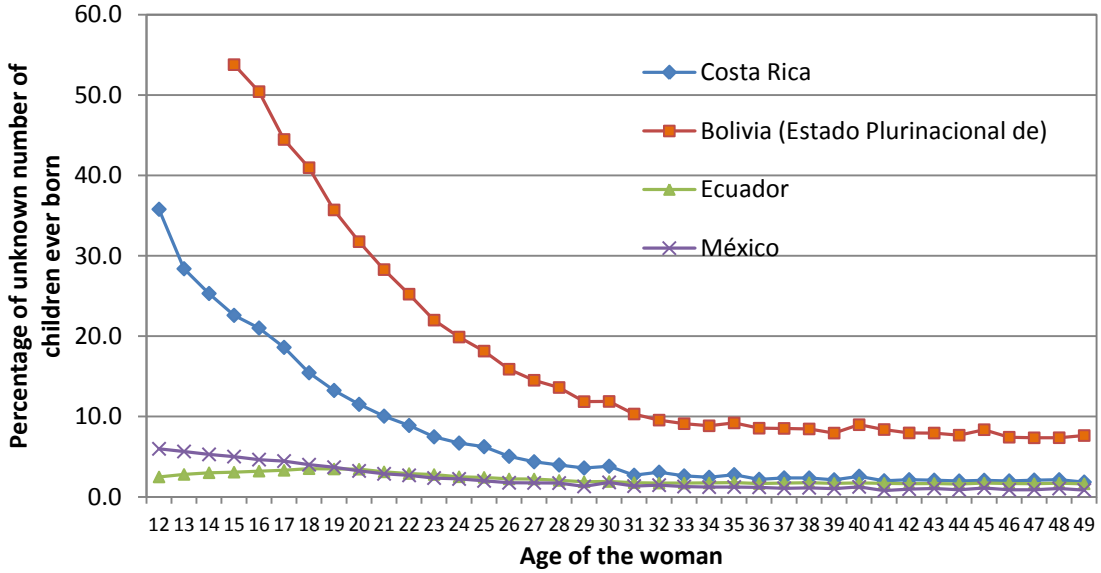
Note [a] Data updated according to Data for 2010 for Chile and Paraguay are preliminary estimates provided by Ruiz, Magda (Cepal/Celade) on April 7, 2015. Other 2010 are data to be presented at the VI ALAP Congress, Lima, Peru, August 12-15, 2015.

It is also important to remember that the censuses are generally 10 years apart, and good estimates of population by age and sex, and for some specific segments, are required for the inter-census period. Without good vital statistics, the uncertainty of the estimates could be very high. Additionally, the panorama of Table 1 is not promising as there is no guarantee that countries will have resources to carry out censuses every 10 years, much less that they will collect all the information needed, because long questionnaires are less and less common. Indeed, due to the large period between censuses, and faced with deficient vital statistics and the lack of migration records, some countries carry out a mid-decade simple

count of the population, as in the case of Brazil. But when confronting economic crises, such counts are among the first items to be cut from the budgets[13].

Besides the coverage of census data, other important problems arise in the quality of the reported information. Graph 1 shows the percentage of the response "unknown" for the number of births per woman in four countries in the 2010 census round. This underscores the diversity of data quality in the region. Two of the countries have a very high percentage of women who did not answer the question on births, and another two have very low percentages. The percentage of "unknown" response decreases with age everywhere, but those in Bolivia and Costa Rica are extremely high for adolescents. In Bolivia, for example, for more than half of the population aged 15 it is unknown whether they are mothers. Another issue that demands notice in Graph 1 is that not all countries have information for women under age 15, and others considered women 12 years or older. As has been pointed out before, the SDG indicator regarding adolescent birth rates for Target 3.7 has to be estimated for women aged 10 to 14. If vital statistics are not available, or there is deficient data, this indicator will not be monitored for these countries. Some could claim to use demographic survey data, however, the sample size of these surveys is not suitable for a quality indicator for an event that is rare in the population at risk.

Graph 1: Percentage of unknown number of births, by women's age, Censuses around 2010



Source: Naciones Unidas, CEPAL/CEPALSTAT. http://interwp.cepal.org/sisgen/MATERNILAC_presentacion.asp, accessed March 27, 2015.

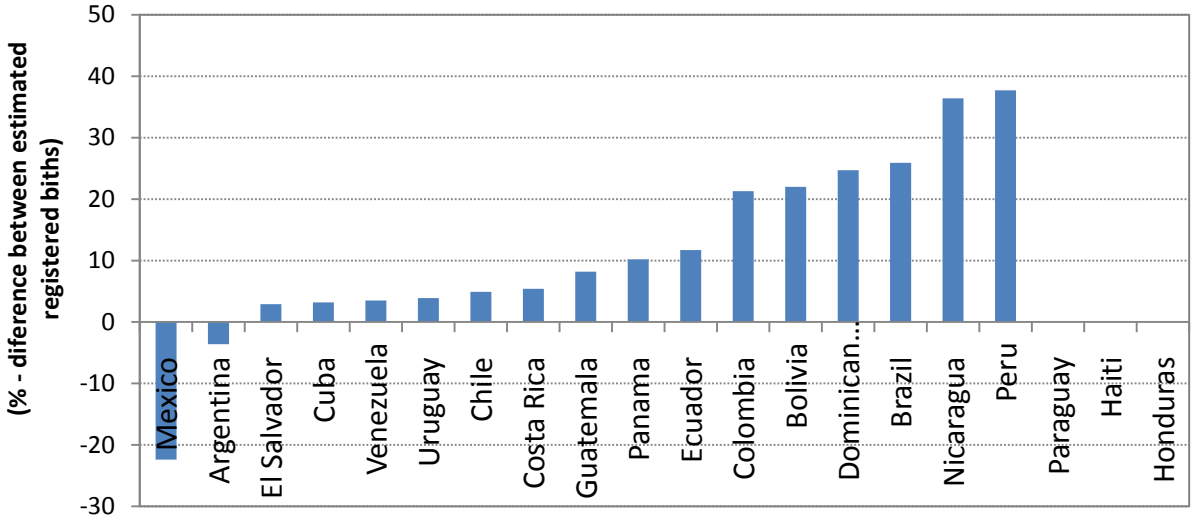
[13] The 2015 population count in Brazil, with data collection having been planned for years, was just cancelled in March 2015.

A second basic and very important source of demographic data, needed for the SDG indicators, as already mentioned, is the registration of vital statistics. These are generally administrative records that come from the registration of births, deaths and marriages. In Latin America the procedures used to produce, manage, and disseminate these data vary from one country to another. A common feature among almost all countries is that this registration does not cover all such events, which leads to regional, socioeconomic and demographic differentials in coverage. Also, having good birth registration does not mean there is also a consistent death registration, or vice-versa.

Graphs 2 and 3 show, respectively, an indicator for the quality of birth and death registration in several countries. Note that coverage of these events varies greatly in the region, from almost zero difference in some countries, like Cuba, to a difference[14] of 40% for births and more than 80% for deaths in El Salvador and Haiti. Also note that at the lower end of the graph some countries do not even present data. It is interesting to observe that the difference between projections and registration is not always positive. In Mexico, Graph 2, for instance, the difference in what could be considered an over-counting of births is higher than 20%, which could be due, for example, to double registration. However, other errors might cause this difference, such as an exaggeration of the fertility rate implicit in the projection or an over-emigration of women of reproductive age (Naciones Unidas, 2014). Indeed, this is a problem that can be reflected in the estimates for other countries also, finding a high difference that could come from overestimated projections.

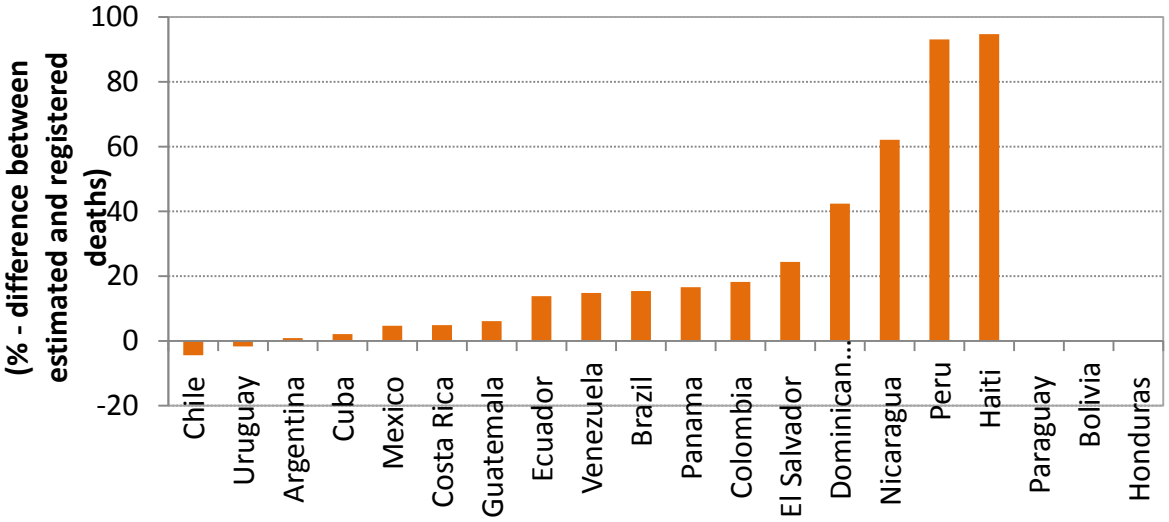
[14] Here it is called difference and not omission because the comparison is made with the rates implicit in the projections, which also are subject to errors.

Graph 2: Relative differences between estimated births by the implicit rates in the population projections and the registered births in the vital statistics registration system, Latin American countries, 2000-2005.



Source: Naciones Unidas (2014).

Graph 3: Relative differences between estimated deaths by the implicit rates in the population projections and the registered deaths in the vital statistics registration system, Latin American countries, 2000-2005.

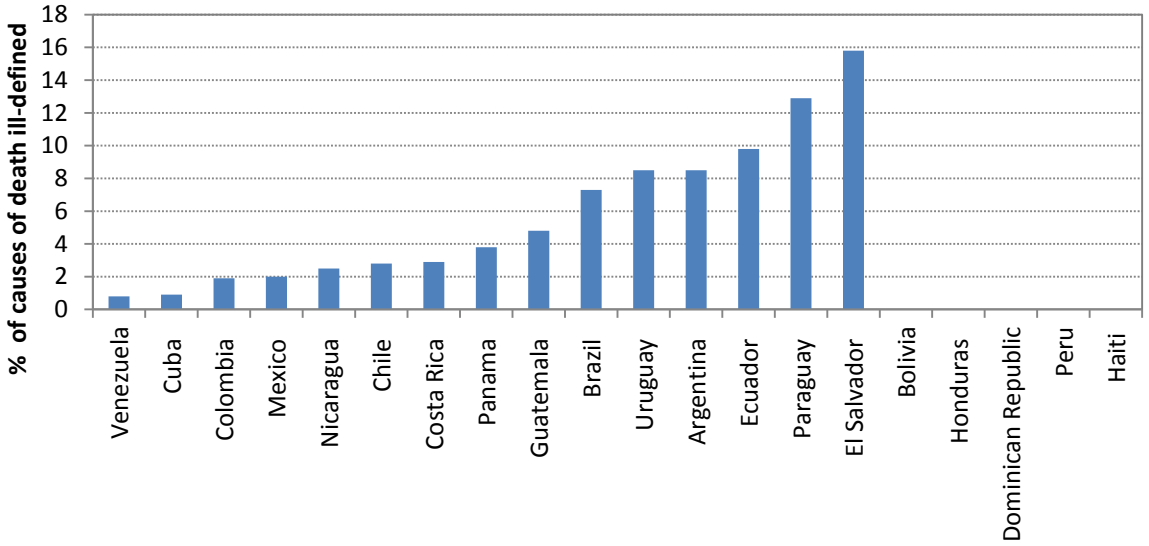


Source: Naciones Unidas (2014).

In addition, in terms of the errors in coverage, the vital statistics system suffers from deficient completion of forms. Here, it is presented simply as one indicator of the quality of registration of deaths, which is the percentage of ill-defined causes of death (Graph 4). This is regularly used as a proxy indicator for registration quality (Naciones Unidas, 2014). The level of poorly

defined causes of deaths varies from close to 1% to around 16% in Latin America. Also of interest is that it is not always the vital statistics systems with best coverage that have 100% completeness. See the examples of Uruguay and Argentina. The information depends on where the deaths occurred, in hospitals or not; on the medical assistance during the process; and on the training the doctors received for completing the form with the correct cause of death.

Graph 4: Percentage of ill-defined causes of death registered in the vital statistics system, Latin American countries, around 2009.



Source: Naciones Unidas (2014).

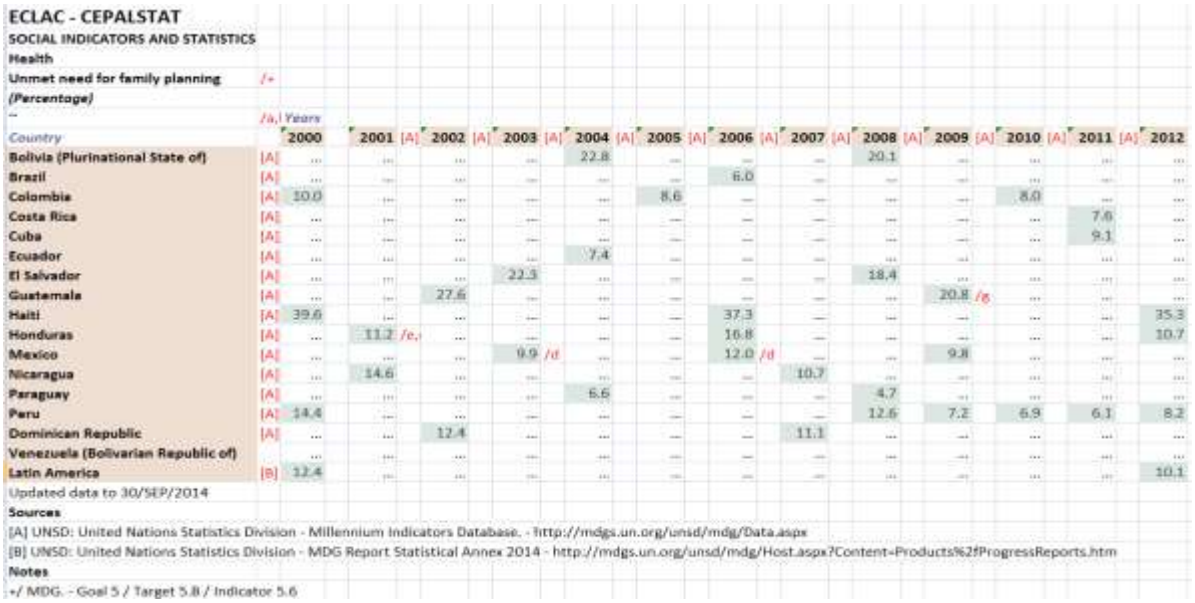
There are several other indicators that demonstrate the defective data on birth and death registration in the region, however, there is another very important source of data that has to be addressed here: the Demographic and Health Surveys (DHS), which collect data for most of the indicators about sexual and reproductive health. In Latin America, some countries have participated since the first wave of DHS, during the 1980's. In total there are 13 Latin American countries that have participated in some wave of the survey, and some during the 1990's (Haiti, Honduras, Nicaragua, Paraguay). Most recently, some countries have carried out their survey independently from Macro International, some with the help and methodology of the U.S. Center for Disease Control (CDC), and some with national resources, like Mexico and Brazil in the 2000 decade.

Even though these surveys are probabilistic and representative of the country, the sample size is too small to allow for certain indicators related to rare events, such as infant mortality

rates, or for disaggregation by age, income, education, etc. Also, the survey is very complex and only experienced data producers achieved good results. Hence, this type of survey is at risk of not being conducted in this region. In Brazil, for example, the last survey conducted with international help was in 1996, and was done by the Ministry of Health in 2006. There has been no updating of the data since then, and there are no plans for another survey.

Figure 1, which is a picture of the output of a table from ECLAC-CEPALSTAT on MDG monitoring indicators, shows the unmet need for family planning from 2000 to the most current data available for Latin American countries. For the vast majority of these countries, the source for data on current use of modern contraception comes from DHS surveys. The picture is not very reassuring for good monitoring of the SDG Target 3.7, which uses this information for one of the indicators selected.

Figure 1: Portrait of the estimates of unmet need for family planning available for Latin American countries, 2000-2012.



ECLAC - CEPALSTAT
SOCIAL INDICATORS AND STATISTICS
Health
Unmet need for family planning /+
(Percentage)
/a,i Years

Country		2000	[A]	2001	[A]	2002	[A]	2003	[A]	2004	[A]	2005	[A]	2006	[A]	2007	[A]	2008	[A]	2009	[A]	2010	[A]	2011	[A]	2012
Bolivia (Plurinational State of)	[A]	...		...		...		...		22.8		...		...		...		20.1		...		...		...		...
Brazil	[A]	...		...		...		...		...		...		6.0		...		...		...		...		...		...
Colombia	[A]	10.0		...		...		...		...		8.6		...		...		...		...		8.0		...		...
Costa Rica	[A]	...		...		...		...		...		...		...		...		...		...		...		7.6		...
Cuba	[A]	...		...		...		...		...		...		...		...		...		...		...		9.1		...
Ecuador	[A]	...		...		...		...		7.4		...		...		...		...		...		...		...		...
El Salvador	[A]	...		...		...		22.3		...		...		...		...		18.4		...		...		...		...
Guatemala	[A]	...		...		27.6		...		...		...		...		...		...		20.8 /6		...		...		...
Haiti	[A]	39.6		...		...		...		...		...		37.3		...		...		...		...		...		35.3
Honduras	[A]	...		11.2 /e,i		...		...		...		...		16.8		...		...		...		...		...		10.7
Mexico	[A]	...		...		...		9.9 /d		...		...		12.0 /d		...		...		9.8		...		...		...
Nicaragua	[A]	...		14.6		...		...		...		...		...		10.7		...		...		...		...		...
Paraguay	[A]	...		...		...		...		6.6		...		...		...		4.7		...		...		...		...
Peru	[A]	14.4		...		...		...		...		...		...		...		12.6		7.2		6.9		6.1		8.2
Dominican Republic	[A]	...		...		12.4		...		...		...		...		11.1		...		...		...		...		...
Venezuela (Bolivarian Republic of)		...		...		...		...		...		...		...		...		...		...		...		...		...
Latin America	[B]	12.4		...		...		...		...		...		...		...		...		...		...		...		10.1

Updated data to 30/SEP/2014
Sources
[A] UNSD: United Nations Statistics Division - Millennium Indicators Database. - http://mdgs.un.org/unsd/mdg/Data.aspx
[B] UNSD: United Nations Statistics Division - MDG Report Statistical Annex 2014 - http://mdgs.un.org/unsd/mdg/Host.aspx?Content=Products%2fProgressReports.htm
Notes
+/ MDG. - Goal 5 / Target 5.B / Indicator 5.6

Source: CEPAL, CEPALSTAT, 2015. Databases and Statistical Publications, available at <http://estadisticas.cepal.org/cepalstat/WEB_CEPALSTAT/Portada.asp?idioma=i>.

Figure 2 shows a similar picture for the maternal mortality ratio. This indicator has been estimated from DHS surveys (and in some countries it is the only source); however, sample size does not allow it to be considered a reliable indicator. This picture, though, seems more encouraging than the previous one because it presents estimates for almost each 50-year period for all countries. Nonetheless, these estimates are available as a result of the Unicef project for monitoring the situation of women and children, and estimates are made from

models and not hard data. The pilot project on maternal mortality carried out by CEPAL/Celade (Cobos, Miller, Ruiz Salgueiro, 2013) identifies some problems regarding the difficulties in estimating maternal mortality in the region and concludes:

> In Latin America, the reasons that underline under-registration and the poor classification of maternal death include the lack of certification or registry in areas inhabited by indigenous populations or in remote zones. The cultural, economic and geographical barriers also play an essential role in this situation. Therefore, to address inequalities and progress towards meeting the MDG it is important to develop policies based on the needs of specific populations. In the most vulnerable sectors, for example, it is necessary to facilitate access to timely diagnosis, assistance in case of gender violence and institutionalizing births.

Figure 2: Portrait of the estimates of maternal mortality ratio available for Latin American countries, 2000-2012.

ECLAC - CEPALSTAT

SOCIAL INDICATORS AND STATISTICS

Health

Maternal mortality ratio /+

(Rate per 100,000 live births)

~ [A] *Years*

Country		1990 /a	1995 /a	2000 /a	2005 /a	2010	2013 /a
Argentina	/a	71	60	63	70	...	69
Bolivia (Plurinational State of)	/a	510	420	330	270	...	200
Brazil	/a	120	100	85	73	...	69
Chile	/a	55	40	29	26	...	22
Colombia	/a	100	81	130	97	...	83
Costa Rica	/a	38	45	44	46	...	38
Cuba	/a	63	60	63	67	...	80
Ecuador	/a	160	130	120	98	...	87
El Salvador	/a	110	96	80	72	...	69
Guatemala	/a	270	220	160	140	...	140
Haiti	/a	670	580	510	470	...	380
Honduras	/a	290	200	150	130	...	120
Mexico	/a	88	77	67	50	...	49
Nicaragua	/a	170	160	140	120	...	100
Panama	/a	98	91	79	83	...	85
Paraguay	/a	130	130	120	130	...	110
Peru	/a	250	220	160	120	...	89
Dominican Republic	/a	240	180	120	130	...	100
Uruguay	/a	42	34	35	32	...	14
Venezuela (Bolivarian Republic of)	/a	93	98	91	94	...	110
Latin America	/a	130	110	98	84	...	77

Updated data to 24/JUN/2014

Sources

[A] UNICEF: United Nations Childrens Fund - Monitoring the Situation of Children and Women - http://www.childinfo.org/

Source: CEPAL, CEPALSTAT, 2015. Databases and Statistical Publications, available at <http://estadisticas.cepal.org/cepalstat/WEB_CEPALSTAT/Portada.asp?idioma=i>.

Finally, another important feature affecting the adequate monitoring of indicators is the time between data production and dissemination. Figure 3 shows adolescent mortality rates from 2000 to 2012, which is the most recent year available at the CEPALSTAT. This indicator for the inter-census period comes from vital statistics or from household surveys. As can be seen in the table, only one country has the data for 2012 and none have data for the following years. Also, some countries have estimates only for long periods of time, which

makes country-specific monitoring very difficult. Meanwhile, the countries that have yearly estimates show instabilities that are probably random and not the real trends of the indicator.

Figure 3: Portrait of the estimates of adolescent[1] birth rate available for Latin American countries, 2000-2012.

ECLAC - CEPALSTAT

SOCIAL INDICATORS AND STATISTICS

Population

Adolescent birth rate /a

(Number of live births per 1,000 adolescent women aged 15-19)

Country		Years 2000		2001	[A]	2002	[A]	2003	[A]	2004	[A]	2005	[A]	2006	[A]	2007	[A]	2008	[A]	2009	[A]	2010	[A]	2011		2012	[A]/a
Argentina	[A]	64.7		61.0		61.5		57.5		63.8		63.8		63.0		64.5		67.5		68.2		69.1		69.6		...	
Bolivia (Plurinational State of)	[A]	...		97.0		...		...		...		89.0		...		...		...		...		...		...		...	
Brazil	[A]	88.1		85.0		81.5		78.9		77.6		76.3		74.5		72.2		70.1		68.0		67.4		64.8		...	
Chile	[A]	60.1		57.7		53.2		48.4		...		48.7		50.4		52.5		53.9		53.6		51.9		...		...	
Colombia	[A]	80.6		...		92.0		...		...		96.2		...		85.0		...		...		...		...		...	
Costa Rica	[A]	83.4		75.4		68.8		69.1		67.7		...		63.1		65.2		69.0		67.2		...		...		...	
Cuba	[A]	49.4		48.5		48.0		46.0		44.9		42.9		40.5		42.9		48.9		50.9		51.4		54.2		...	
Ecuador	[A]	...		100.0		99.6		...		...		...		...		...		...		...		...		...		...	
El Salvador	[A]	98.7		88.5		80.6		75.9		...		69.6		64.4		64.7		...		...		61.4		63.3		...	
Guatemala	[A]	120.0	/b	...		...		99.8	/b	99.9	/a	...		98.0	/b	...		...		...		88.0	/a	92.4	/a	...	
Haiti	[A]	...		...		...		69.0		...		...		...		...		...		65.0		...		...		...	
Honduras	[A]	...		...		...		108.0		...		...		...		...		...		99.0		...		...		...	
Mexico	[A]	87.8		87.1		84.3		80.4		80.5		78.8		75.6		82.3		83.7		85.9		87.7		87.1		84.6	
Nicaragua	[A]	...		...		...		...		...		108.5		...		...		...		...		92.0		...		...	
Panama	[A]	85.5	/a	89.3	/b	77.7	/a	77.3	/a	84.9	/b	77.8	/a	78.5	/a	80.1	/a	82.8	/a	84.1	/a	81.4	/a	...		...	
Paraguay	[A]	...		64.0		...		...		...		...		...		63.0		...		...		...		...		...	
Peru	[A]	...		59.0		61.0		...		...		64.0		...		69.0		72.0		62.0		67.0		...		...	
Dominican Republic	[A]	118.0		...		86.4		...		98.0		...		...		96.2		...		...		...		...		...	
Uruguay	[A]	66.9		68.4		67.2		62.0		60.3		61.5		62.0		59.6		...		...		...		...		...	
Venezuela (Bolivarian Republic	[A]	91.7	/a	86.4	/a	80.2	/a	92.1	/b	91.7	/b	108.9	/a	106.3	/a	100.7	/a	...		98.6	/a	98.1	/a	101.2	/a	...	
Latin America	[B]	86.7		...		...		...		...		...		...		...		...		...		...		76.4		...	

Updated data to 08/SEP/2014

Sources

[A] UNSD: United Nations Statistics Division - Millennium Indicators Database. - http://mdgs.un.org/unsd/mdg/Data.aspx

Source: CEPAL, CEPALSTAT, 2015. Databases and Statistical Publications, available at <http://estadisticas.cepal.org/cepalstat/WEB_CEPALSTAT/Portada.asp?idioma=i>.
Note 1 - Adolescent includes women aged 15-19 years old.

Measures needed to produce indicators that inform public policies

A data revolution in the region is greatly needed. Not quite yet a data revolution that relies on the unstructured data coming from the "internet of things". Latin America is not only ready to engage in the endeavor of a data revolution, but it is essential if the region is to yield adequate hard data for obtaining the minimum information about population dynamics, reproductive health, among others known to be produced in the older existing datasets. Governments should invest seriously in human and financial resources to produce the data that will guide the future agenda. This process is not to be done flying solo; on the contrary, the advances in technology are best utilized in partnership, and based on the experiences of those countries and professionals that have had good results.

Another very important process to be taken seriously in each country is setting up a well functioning national statistical system (NSS), based on the UN recommendations. Given the variety of data producers and data sources, if there is not an effective and respected coordination of the NSS, it could lead to a waste of resources (humans and financial) and the appearance of competing estimates of the indicators. Also, this coordinating group must establish methodologies and processes that allow for data integration, and linkage of data, to increase the reach of the data thematically and longitudinally.

Lastly, but not least, an important step to take in the region is to improve training in demography. "*The International Union for the Scientific Study of Population has been concerned for some time at the erosion of demographic skills and knowledge within National Statistical Offices*" (IUSSP, 2014, p. 3), There is no single or simple answer to address this problem. There are different trainings needed, both short and long term. The long term involves a Masters and PhD level of education, and not all statistical offices have planned training for employees, especially if they have an agenda of data publication and a shortage of human resources. Nonetheless, this issue should be given high priority.

References

A WORLD THAT COUNTS: Mobilising the Data Revolution for Sustainable Development (2014). Report, Independent Expert Advisory Group on a Data Revolution, Nov. 2014. Available at <http://www.undatarevolution.org/wp-content/uploads/2014/12/A-World-That-Counts2.pdf>. Accessed March 22, 2015.

Doran, G. T. (1981). "There's a S.M.A.R.T. Way to Write Management's Goals and Objectives", Management Review, Vol. 70, Issue 11, pp. 35-36.

Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage. Available at <http://books.google.com.br/books?hl=pt-BR&lr=&id=rNjSAwAAQBAJ&oi=fnd&pg=PT10&dq=%22data+revolution%22&ots=tCPwlgjOvn&sig=XOkvGT9pYFr0W7MQmPzaSEKeU4k#v=onepage&q=%22data%20revolution%22&f=false>

Cobos, M. I., Miller, T., Ruiz Salgueiro, M. Hacia la armonización de las estimaciones de mortalidad materna en América Latina: Hallazgos de un estudio piloto en ocho países. Naciones Unidas, Santiago: ECLAC, 2013.

Friedman, J. H. (2001). The role of statistics in the data revolution?. *International Statistical Review*, 69(1), 5-10.

META Group. "3D Data Management: Controlling Data Volume, Velocity, and Variety." February 2001.

Naciones Unidas (2013), Principales cambios en las boletas de los censos latinoamericanos de las décadas de 1990, 2000 y 2010, **Serie Manuales,** Celade/Cepal: Santiago de Chile, noviembre 2013.

Naciones Unidas (2014), Los datos demográficos: Alcances, limitaciones y métodos de evaluación, **Serie Manuales**. Celade/Cepal: Santiago de Chile, octubre de 2014.