Data Note - ID4D Global Dataset and ID4D-Findex Survey

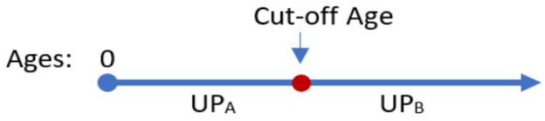
ID4D Global Dataset

How is the number of people without an official proof of identity calculated globally and on the country/economy level?

The ID4D Global Dataset uses a combination of self-reported data from ID-issuing authorities as well as other publicly available data - such as UNICEF birth registration and voter registration rates - to produce a global estimate of the ID gap. For 2018, this estimate stands at 1 billion based on data from 151 economies.[1]

Measuring the identity gap is a challenge given the lack of an internationally-agreed upon definition of "proof of legal identity" and the fact that individuals are likely to have different identity credentials throughout their lifetime. In most countries, for example, children are primarily identified through birth registration, while teenagers and adults may have other forms of ID, such as national IDs and voter registration. To address this issue, the Dataset uses different metrics to estimate the unregistered population (UP) for those under and over a cut-off age, which varies by country. Birth registration is used to estimate the unregistered population below the cut-off age (UP_A ~ children); direct administrative data or voter data are used to estimate unregistered population above (and including) the cut-off age (UP_B ~ adults). These estimates are then added together to produce the total UP estimate. This approach is illustrated below:

$$UP = UP_A + UP_B$$



Most birth registration rates are drawn from UNICEF's global database, which compiles birth registration data from nationally representative surveys (e.g., DHS and MICS) and population censuses. In some cases, this data only provides estimated registration rates for children under 5, but not older children. In such cases, it is assumed that the birth registration rate is constant for age groups from 5 through to the cut-off age (UP_A).

To calculate the unregistered population above the cut-off age (UP_B), direct administrative data is used to the extent possible, which may be directly reported by the national ID entity or obtained from the official website of national ID entity or from other, official sources (e.g. from a survey or census). When no direct data are available, the most recent voter registration data is used as a proxy indicator[2], with the key assumption that registering to vote requires some form of proof of legal identity. Voter registration data was retrieved from national electoral commissions as well as databases by the International Institute for Democracy and Electoral Assistance (IDEA) and the International Foundation for Electoral Systems (IFES).

The cut-off age depends on the source of the data. When voter registration data is used, cut-off age equals the voting age, which is typically, though not always, 18 years old. When direct

[1] Economies without national ID program and with an under-5 birth registration rate of over 95% are excluded; high income countries with an under-5 birth registration rate of over 99.9% are excluded; China is excluded as there is no publicly available data.

[2] Voter registration data is used as a proxy indicator for the ID coverage of adults in 120 economies where ID coverage data from official sources was not available.

administrative ID coverage data is used, the cut-off age equals the mandatory (national) ID age (typically 16 or 18 years old). Where the mandatory ID age is above the minimum ID age and/or when no age-disaggregated data is available from ID authorities starting at the mandatory ID age, the cut-off age is adjusted to be the lowest age captured in the data. The denominators for calculating the percentage of the unregistered population are the 2018 population estimates from the UN World Population Prospects (2017 edition) unless otherwise noted (see footnote 3).

What are the main limitations of this data?

- The Dataset was created to measure the scale of the *overall global identification gap*; estimates for individual economies are subject to considerable uncertainty and should NOT be used as precise point estimates.
- Not all children whose births have been registered will have been issued a birth certificate (Bhatia et al. 2017). In addition, UNICEF (2013) notes that since birth registration data is mostly survey-based, respondents may not always be aware of formal birth registration and certification processes, resulting in erroneous data. Consequently, the Global Dataset may *underestimate* the number and share of people without a proof of legal identity.
- Some countries allow all residents—including non-nationals—to register in their ID systems, while others are limited to citizens only. Residents excluded from a country's foundational (national) ID system may have other official IDs, such as passports issued by their country of origin, or may not have access to any government-recognized, widely accepted IDs. The Dataset cannot provide information about ID coverage gaps for marginalized populations, such as migrants, refugees, and stateless persons.
- Where voter registration data is used as a proxy to estimate the number of people without a proof of identity, persons who choose not to or are unable to register to vote (e.g. non-nationals) will be excluded, which could result in an *overestimate* of the number of unregistered people[3]. At the same time, registering to vote or obtaining a voter credential may not always require or produce a trusted, widely accepted identity credential, in which case the effective identification gap may be *larger* than suggested by the Dataset.
- Registration data—whether obtained directly from ID authorities or based on voter registration—may exceed the total population size estimate for the relevant age group, resulting in 'excess registration'[4]. In these cases, the unregistered population size for those above the cut-off/ID age is set to zero to avoid negative UP estimates. For countries where gender disaggregated data is available and it shows a positive number of unregistered people among one of the genders (suggesting a coverage gap in one gender group and 'excess' registration in the other), the overall UP_B is set to equal the unregistered number of that gender.

[3] This can pose a particularly large challenge for economies with a large foreign population share. For Oman, Qatar, Saudi Arabia, and United Arab Emirates, where voter registration was used a proxy indicator, total population was adjusted to exclude foreign residents, using data from the Gulf Labour Markets, Migration, and Population Programme of the Gulf Research Center and the Migration Policy Centre of the European University Institute.

[4] 'Excess registration' may be due to several reasons, such as, inaccurate population estimates if census data are unreliable or several years old; a (large) share of the country's registered population living abroad; the existence of duplicate records in the registry; or non-removal of deceased persons from the registry.

- Although ID4D makes every effort to use the most up-to-date information, the data on which the country estimates for the unregistered population rely may be several years old (e.g. surveys measuring birth registration and/or the latest voter registration effort can go back several years).

ID4D-Findex Survey Data

How was the survey data collected?

ID4D partnered with the World Bank's Global Findex team to gather additional data on ID coverage, ID use, and barriers to obtaining a proof of identity. The Findex survey covers over 100,000 people in 99 economies—representing 74 percent of the world's population. The survey was carried out over the 2017 calendar year by Gallup, Inc., as part of its Gallup World Poll. Approximately 1,000 people were surveyed in each economy, using randomly selected, nationally representative samples of the non-institutionalized population age 15 and above. For more details on the survey methodology, please see the Global Findex methodology note.

How is the share and number of people without an ID measured?

To estimate ID coverage in each economy, respondents were asked whether they personally had a specific foundational/national ID, using local terminology to the extent possible (e.g. 'Kartu Tanda Penduduk' for Indonesia or 'Aadhaar' for India). The exact terminology used for the survey in each economy can be found here. To estimate the number of people above the age of 15 without proof of identity, the share of respondents who responded "no" to having an ID was multiplied by the 2018 population estimate for the economy's 15-and-above age group from the UN World Population Prospects (2017 edition).

What are the key limitations of this data?

- The Findex data provide an estimate of ID coverage rates for those over the age of 15, however they do not include children and are thus not representative of the entire population.
- Although the survey almost always used specific terminology for the country's foundational/national ID, respondents may have interpreted the question to also apply to other widely-held identity documents, such as voter ID cards or other functional IDs (e.g. social/health service ID, driver's license, passport).
- Survey respondents may also feel inclined to respond affirmatively about having an ID document, even when this may not be the case, due to social desirability bias or an (unfounded) fear of some negative consequence, which would result in an *underestimate* of the unregistered population.
- Since the survey includes respondents down to the age of 15, the Findex data may *overestimate* the ID coverage gap by a few percentage points in countries where IDs are issued at a later age (see the ID4D Global Dataset for a full list), since some of the youngest respondents would not have been able to legally obtain the specified ID.
- In a few countries (e.g. Armenia, Tajikistan), survey respondents were asked about the national ID, whereas other identity documents, for instance internal passports, are more widely held among the population and used for a variety of purposes. In such cases, the Findex data may *overestimate* the unregistered population in the given economy.

- An economy's ID coverage can change rapidly as it is possible to enroll millions of people in an ID system within a matter of months. A few countries (e.g. Malawi) have made significant progress in their registration efforts since the surveys were concluded in 2017.

Why does the share/number of the unregistered population differ between these two data sources?

There is currently no internationally-agreed or standard approach to define and measure the number of people without proof of legal identity and neither dataset provide a perfect measure at the country/economy level. These two datasets help triangulate the size of the ID gap by using different sources and methodology. Together, they provide a reasonable estimate about the overall scale of the global identification gap and can help highlight which regions or groups of economies are in greatest need of assistance. However, estimates of the unregistered population for individual economies can differ substantially between the two datasets due to the different data collection methods, the different indicators used to measure ID coverage, and the different populations they cover. The key differences between the two datasets are summarized below:

	ID4D Global Dataset	ID4D-Findex Survey Data
Data source(s)	• Above cut-off age (~adult): Administrative data: • Below cut-off age (~child): Survey-based birth registration data (UNICEF)	Nationally representative survey, collected as part of the Gallup World Poll

Type of indicator used to estimate coverage	• Above cut-off age (~adults):
	  • **National/foundational ID registration data from ID authorities,** where available; OR
	  • **Voter registration** used as a proxy for having a proof of identity in 120 economies
	• Below cut-off age (~children): **Under-5 birth registration**
		Survey respondents were asked whether they personally had a **specific foundational/national ID** (birth certificates, voter IDs, etc. not considered)

Population covered	Global and economy-level estimates include **all people aged 0 and above**; global estimate based on 151 economies	ID coverage data limited to people **aged 15 and above;** data available for 99 economies