Reducing Inequality in FEALAC Member Countries

Methodology

STRENGTHENING THE EVIDENCE BASE TO LEAVE NO ONE BEHIND

United Nations
ECLAC

Forum for East Asia-Latin America Cooperation

UNITED NATIONS
ESCAP
Economic and Social Commission for Asia and the Pacific

ST/ESCAP/2939

Disclaimer

The views expressed in this publication are those of the authors and do not necessarily reflect the views and policies of the United Nations or other international agencies. The publication has been issued without formal editing. Reproduction and dissemination of material in this publication for educational or other non-commercial purposes are authorized without prior written permission from the copyright holder, provided that the source is fully acknowledged. For further information on this publication, please contact:

Social Development Division
Economic and Social Commission for Asia and the Pacific
United Nations Building
Rajadamnern Nok Avenue
Bangkok 10200, Thailand
Email: escap-sdd@un.org
Website: www.unescap.org

Reducing Inequality in FEALAC Member Countries

Methodology

STRENGTHENING THE EVIDENCE BASE TO LEAVE NO ONE BEHIND

UNITED NATIONS ESCAP
Economic and Social Commission for Asia and the Pacific

UNITED NATIONS
ECLAC

Forum for East Asia-Latin America Cooperation

Table of contents

List of figures

List of tables

1 Introduction

Over the past decades, both the Asia-Pacific and Latin America and the Caribbean regions have made considerable strides in social development driven by economic growth, generating new jobs, increasing labour incomes, strengthening social protection systems, and improving access to basic services and other opportunities. Despite this sustained economic development and substantial reductions in poverty, very large inequalities continue to exist on the basis of wealth, gender, residence, level of education, among other factors.

High levels of inequality not only stifle economic progress, but also negatively affect feelings of trust and social cohesion, posing a formidable barrier to sustainable development (ESCAP, 2017). This inequality within countries has sparked public concern and academic interest. The 2030 Agenda for Sustainable Development has a stand-alone goal on inequality: Sustainable Development Goal (SDG) 10 to "reduce inequalities within and among countries" is thus a core policy priority to ensure a sustainable and prosperous future for all.

2 Leaving No One Behind in the context of the 2030 Agenda

To reduce inequalities, the 2030 Agenda takes an ambitious yet pragmatic approach, stressing that no one should be left behind in any of its Goals — and that the furthest behind should become the focus of policymaking. It states:

"As we embark on this great collective journey, we pledge that no one will be left behind. Recognizing that the dignity of the human person is fundamental, we wish to see the Goals and targets met for all nations and peoples and for all segments of society. And we will endeavour to reach the furthest behind first." (UN, 2015, paragraph 4)

Member States have explicitly called on the United Nations and its agencies, funds and programmes to implement the Leave No One Behind (LNOB) pledge. The United Nations system has responded promptly, bringing the LNOB pledge at the core of its programming, beginning with the necessary disaggregation of data and statistics. SDG target 17.18 calls to "enhance capacity building support to developing countries, including for LDCs and SIDS, to increase significantly the availability of high-quality, timely and reliable data disaggregated by income, gender, age, race, ethnicity, migratory status, disability, geographic location and other characteristics relevant in national contexts" by 2020. Consequently, the United Nations System Shared Framework for Action calls for "greater data disaggregation across a wider range of grounds for all SDG indicators; systematic analysis of available (disaggregated) data on marginalized groups; new tools for analysing horizontal and vertical inequalities, as well as discrimination, stigma, exclusion, and equity issues; identification of subjects of multiple and intersecting forms of discrimination; joined-up analysis of the drivers, root causes and underlying determinants of inequalities and discrimination" (UN, 2017a).

Further, the UN Sustainable Development Cooperation Framework (previously the UN Development Assistance Framework) places the pledge to leave no one behind at the core of its four principles for unifying programming and advocacy, requiring all UN entities "to prioritize [their] programmatic interventions to address the situation of those most marginalized, discriminated against and excluded, and to empower them as active agents of development" (UN, 2017b). The methodology presented here is thus of direct use for generating discussions on this topic, corresponding to Steps 1, 4 and 5 of the five-step methodology developed by the UNSDG Operational Guide for UN Country Teams, assisting Member States in operationalizing the pledge to LNOB and reach the furthest behind first (UNSDG, 2019).

3 The data

In practice, LNOB means moving beyond assessing average and aggregate progress, towards ensuring progress for all population groups at a disaggregated level. This requires disaggregating data to identify groups being excluded or discriminated against, as well those experiencing multiple and intersecting forms of discrimination and inequalities.

The methodological tool presented in this paper (Classification and Regression Tree (CART) Analysis) uses the Demographic and Health Surveys (DHS) and the Multiple Indicator Cluster Surveys (MICS). DHS and MICS are publicly available for 8 Asian and Pacific countries and 15 countries in Latin America and the Caribbean in the FEALAC region. The datasets are selected because of a) comparability across countries; b) accessibility of the data; and c) the rich set of questions on health, demographic and basic socioeconomic data that refer both to the household (e.g. water and sanitation, financial inclusion, electricity and clean fuels, as well as ownership of mobile phones/bank cards) and to individuals (e.g. level of education, nutrition status, access to basic healthcare services for women.) Multiple countries have surveys representing two different points in time. The full list of countries and the latest survey years is provided in Table 1.

Despite their many advantages, DHS and MICS also have shortcomings. For example, because some questions are answered at the household level, they do not allow for calculation of sex-disaggregated data. Furthermore, men are not always asked the same sets of questions as women. Lastly, the surveys do not capture people least likely to be counted and reflected in national statistics, such as the homeless, slum dwellers, irregular migrants, nomadic or displaced populations, stateless persons, criminalized populations (e.g., people who use drugs, sex workers) and people in temporary shelters or institutions.

TABLE 1

DHS and MICS availability in Latin America and the Caribbean and Asia and the Pacific in the FEALAC region

1a. Latin America and the Caribbean

COUNTRY	MOST RECENT YEAR AVAILABLE	SURVEY
Argentina	2012	MICS
Bolivia (Plurinational State of)	2008	DHS
Colombia	2015	DHS
Costa Rica	2011	MICS
Cuba	2014	MICS
Dominican Republic	2013	DHS
El Salvador	2014	MICS
Guatemala	2014–2015	DHS
Honduras	2011–2012	DHS
Mexico	2015	MICS
Panama	2013	MICS
Paraguay	2016	MICS
Peru	2012	DHS
Suriname	2010	MICS
Uruguay	2013	MICS

1b. Asia and the Pacific

COUNTRY	MOST RECENT YEAR AVAILABLE	SURVEY
Cambodia	2014	DHS
Indonesia	2017	DHS
Lao People's Democratic Republic	2017	MICS
Mongolia	2018	MICS
Myanmar	2016	DHS
Philippines	2017	DHS
Thailand	2019	MICS
Viet Nam	2013	MICS

Source: ECLAC and ESCAP elaboration based on https://dhsprogram.com/ and https://mics.unicef.org/.

4 The indicators

The indicators used in the analysis cover secondary education, stunting among children under 5 years of age, skilled birth attendance during childbirth, basic drinking water, and clean fuels. The connection between related indicators and the Sustainable Development Goals (SDGs) was the main criterion for their selection[1] (Table 2). Additional indicators are also explored in a dedicated electronic database. The exact questions from DHS and MICS questionnaires with a brief description can be found in Table 3.

TABLE 2
Indicators used in analysis and related SDG indicators

INDICATORS	CLOSEST SDG INDICATOR REFERENCE
Secondary education	4.1.1 Proportion of children and young people: (a) in grades 2/3; (b) at the end of primary; and (c) at the end of lower secondary achieving at least a minimum proficiency level in (i) reading and (ii) mathematics, by sex
Stunting in children under 5 years of age	2.2.1 Prevalence of stunting (height for age <-2 standard deviation from the median of the World Health Organization (WHO) Child Growth Standards) among children under 5 years of age
Skilled birth attendance during childbirth	3.1.2 Proportion of births attended by skilled health personnel
Basic drinking water	6.1.1 Proportion of population using safely managed drinking water services
Clean fuels	7.1.2 Proportion of population with primary reliance on clean fuels and technology

Source: ESCAP elaboration.

TABLE 3
How indicators are identified and defined in DHS and MICS

INDICATOR	VARIABLE NAME	SURVEY QUESTION (IN DHS/ MICS)	DESCRIPTION	SURVEY RECODE
Secondary education	DHS: HV109 MICS: ED4A, ED4B	What is the highest level of school you attended: primary, secondary, or higher?		PR
Stunting in children under 5 years of age	DHS: HC70 MICS: HAZ	Height in centimeters for children age 0–5	If the height of the child is two standard deviations below the average of children of the same age, he/she is considered stunted	PR
Skilled birth attendance during childbirth	DHS: M3A, M3B, M3C, M3D, M3E, M3F MICS: MN17A, MN17B, MN17C, MN17D, MN17E, MN17I, MN17J, MN17K	Who assisted with the delivery of (name)?	Skilled birth attendance includes doctor, nurse, and midwife	IR
Basic drinking water[a]	DHS: HV201 MICS (4&5): WS1	What is the main source of drinking water for members of your household?	Population using improved drinking water sources such as piped household water connection, public standpipe, borehole, protected dug well, protected spring, rainwater collection	HH
Clean fuel	DHS: HV226 MICS: HC6	What type of fuel/energy does your household mainly use for cooking?	Clean fuel includes natural fuel (e.g. compressed natural gas or liquified petroleum gas) or a blend (e.g. gasohol) used as a substitute for fossil fuels and which produces less pollution than the alternatives	HH

Source: ESCAP elaboration.
Note: PR=household member recode; IR= individual recode; HH=household recode.
a Instead of "safely managed", using the "basic services" definition so as to cover more countries.

1 The latest indicators to be used for monitoring the SDGs can be found at: https://unstats.un.org/sdgs/iaeg-sdgs/.

5 The determinant factors (circumstances)

The approach proposed is to identify a set of desired outcomes or opportunities and to measure the gaps among groups in these areas. To do so, a set of "circumstances" is selected from available variables in the DHS and MICS datasets to define the groups. These circumstances are usually a set of conditions that the individuals or the households have little control over.

The selection of variables is consistent across all surveys to maintain comparability of inequality across countries. Ultimately, these circumstances (determinant factors) define the composition of the groups. However, circumstances should not be interpreted as "causes" of inequality. Furthermore, there are many other factors that these models cannot consider, given the limited variables available in the datasets.

Ideally, it would have been preferred to include only circumstances over which a household member had almost no control, such as dominant religion in a household where a respondent is born, ethnicity, existence of a disability, education of the mother or father of the respondent. The majority of the DHS did not include these questions. Some MICS, however, did ask questions related to ethnicity, language and religion.[2] In the cases where these questions were included, the analysis can be repeated using these additional determinant factors. Additional potentially useful factors that could have been of interest for the study are geographical variables, such as province or region of a given country, but that would have affected comparability across countries. These geographic variables are analysed in the work that focuses on one country only.

TABLE 4
Circumstances used to determine groups, per indicator

INDICATORS			CIRCUMSTANCES USED TO DETERMINE GROUPS OF THE FURTHEST BEHIND/AHEAD						
NO.	INDICATORS	REFERENCE GROUP IN SURVEY	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5	FACTOR 6	FACTOR 7
1	Secondary education	Household member aged 20–35[a]	Wealth	Residence (rural or urban)	n/a	Woman/ Man	n/a	n/a	n/a
2	Stunting in children under 5 years of age	Children aged 0–5 who have been measured	Wealth	Residence	Mother's education	Boy/Girl	Number of children <5 years of age	n/a	n/a
3	Skilled birth attendance during childbirth	Women aged 15–49 who have given birth in the past 3-5 years	Wealth	Residence	Responder's education	n/a	Number of children <5 years of age	Age: 15–24, 25–34, 35–49	Single, currently/ formerly married or in union
4	Basic drinking water	All households	Wealth	Residence	Highest education in household	n/a	n/a	n/a	n/a
5	Clean fuels	All households	Wealth	Residence	Highest education in household	n/a	n/a	n/a	n/a

Source: ESCAP elaboration.

a For assessing completion of secondary education, the sample has been restricted to those aged 20–35 (for secondary education) and aged 25–35 (for higher education). The reason is to avoid: (1) skewing the results because of an older population with significantly lower education levels; and (2) including individuals that, because of their young age, could not have completed their education.

2 More recent MICS and DHS datasets started including more questions on disability, migration status, etc.

6 The Classification and Regression Tree (CART) methodology

Knowing that inequality is broadly associated with specific circumstances opens the door to a more in-depth exploration of the data to see exactly which groups are the most marginalized and which groups have benefitted most from development. Identifying these groups could help policymakers better focus policy and programmes to identify the furthest behind.

The primary goal of using the Classification and Regression Tree (CART) Analysis is therefore to identify the groups with the lowest and highest levels of access to opportunities or outcomes, using the selected indicators. The CART Analysis allows for this identification to happen quickly, accurately and based on evidence, while presenting the results in a visually intuitive tree representation (ESCAP, 2020). The indicators used are the "response variables", while the factors that characterize these groups are defined as "circumstances" (independent variables). A classification tree is an analytical structure that represents groups of the sample population that have significantly different response values, or different levels of access.

A classification tree is constructed for each country, using R, an open source statistical software. The root node of the tree is the entire population sample. The tree method algorithm starts by searching for the first split (or branch) of the tree. It does so by looking at each circumstance and separating the sample in two groups, so that it achieves the most "information gain". This information metric can be defined in a few ways, while the most common one – and the one used in this analysis is the "entropy" (Kelleher, Mac Namee and D'Arcy, 2015).

For example, the algorithm estimates access to skilled birth attendance during childbirth by partitioning women into different groups based on the individual circumstances chosen. The formula that represents the core of the algorithm is the following:

$$p(Y_i = 1 || X_{1i}, X_{2i}, \dots, X_{li}) = \sum_{j=1}^{m} p_j \times I((X_{1i}, X_{2i}, \dots, X_{li}) \in A_j)$$

where Y_i is the observed indicator for the i-th individual in the sample, and X_{1i}, ..., X_{li} are the circumstances for the individual. In the example of access to skilled birth attendance during childbirth, Y is the rate of skilled birth attendance during childbirth, X_1, X_2, X_3, X_4, X_5, X_6 (where l = 6) are residence, household wealth level, education, marital status, age group, and number of children in the household, and p represents the probability of having access to skilled birth attendance.

A_1, A_2, ... , A_m are the different partitions of the sample, also called end nodes, where:

$$A_i \cap A_j = \emptyset$$

and

$$\cup_{i=1}^{m} A_i = \Omega$$

This means the end nodes are mutually exclusive and complementary, and every woman who has recently given birth belongs to one and only one of the end nodes. I () only takes value 1 when the i-th household belongs to j-th end node, otherwise, I () takes value 0. The tree algorithm generates the end nodes, according to metrics that measure the effectiveness of the partition that gives to different levels of access to skilled birth attendance during childbirth.

Information theory and entropy is a very common choice for the metrics. Entropy for j-th end node can be calculated according to the definition:

$$I_E(p_j) = -\left(p_j \times log_2 p_j + (1-p_j) \times log_2(1-p_j)\right)$$

The aggregated entropy for the tree is calculated by:

$$H(T) = \sum_{j=1}^{m} q_j \times I_E(p_j)$$

where q_j is the sample proportion of A_j. The actual algorithm that generates the end-nodes is step-by-step, starting from the entire sample. Each time the sample is partitioned new end-nodes are generated and the entropy is calculated and compared to the entropy before the new partition. Each partition (and hence the new end nodes) is kept when the reduction of entropy is bigger than a pre-set threshold. The algorithm stops when no more information gain can be made by new partition, or a set of pre-set conditions cannot be satisfied.

CART also operates under the limitation that each group should have enough group members. To avoid a too small sub-sample size, the analysis has set the tree nodes to have a minimum size of at least 10 per cent of the total population and the split of tree is only made when an information gain criterion is satisfied.

A CART Example

To illustrate how the classification tree identifies the most disadvantaged or advantaged groups, the example of access to skilled birth attendance during childbirth in Lao People's Democratic Republic is used.

Indicator ("response variable"): Access to skilled birth attendance during childbirth in Lao People's Democratic Republic

Factors ("circumstances"): The circumstances being considered are the following:

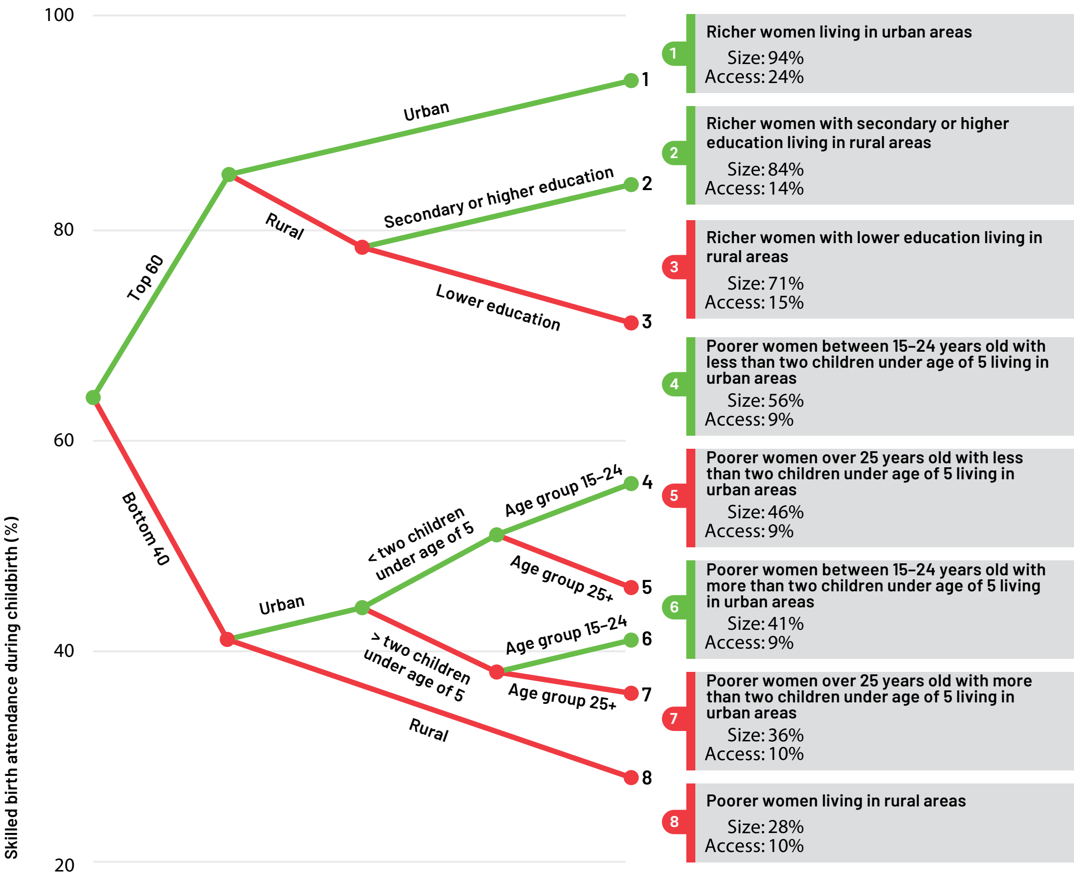
1. Household wealth (Bottom 40 or Top 60),
2. Education (Secondary or Higher vs. Primary or None),
3. Number of children under 5 years of age,
4. Residence (Rural or Urban),
5. Age group (15–24, 25–35, 35+),
6. Marital status (Single, Currently/Formerly Married or in Union).

The classification tree starts at the average access rate of 64 per cent. The algorithm determines that the first split into branches is wealth, specifically where in the wealth distribution a woman belongs: the top 60 per cent or the bottom 40 per cent. Women belonging to the top 60 per cent group have 85 per cent access rate to skilled birth attendance during childbirth, compared with only 41 per cent for those in the bottom 40 group.

In the same example, the algorithm determines a second split for the less advantaged (bottom 40 group) based on their place of residence. Women living in urban areas have an access rate of 44 per cent. That rate falls to 3 in 10 for women living in rural areas, who constitute the furthest behind group. The rate of access to skilled birth attendance also varies among women with children under 5 years of age in urban areas: half of the women with fewer than two children under 5 years of age get skilled birth attendance during childbirth, while 4 in 10 of those with more than two children under 5 years of age do. Age determines the final split for both groups, where women older than 25 years old have a lower access to skilled birth attendance during childbirth when compared to their younger counterparts between 15 and 24 years old. Among the women belonging to the top 60 group, residence constitute the first split, while the top 60 per cent women living in rural areas is further disaggregated by their educational level. Women belonging to the top 60 per cent of the wealth distribution living in urban areas have the highest rate to skilled birth attendance during childbirth, at 94 per cent, being the best-off group.

FIGURE 1

Classification tree highlighting differences in women's access to skilled birth attendance during childbirth in Lao People's Democratic Republic, 2017 (women 15–49 years of age)



Source: ESCAP calculations, using data from the latest DHS and MICS surveys for countries in the Asia-Pacific region.

7 Gaps and limitations

The uniqueness of the Classification and Regression Tree (CART) approach is that it becomes very clear where policies should, or should not, be focused to reach those furthest behind first.[3] The furthest behind groups are identified based on their shared circumstances, thus revealing which socioeconomic features are associated with the biggest gaps in access to basic opportunities. The identity of these groups, their shared circumstances, also point to intersections of disadvantage, since groups usually share more than one circumstance: the furthest behind women in the previous example are not only in the bottom 40 of the wealth distribution, but also live in rural areas. By highlighting such intersectionalities, policy can be better tailored to the furthest behind groups.

The methodology presented in this paper has several advantages, but also some limitations. Firstly, the available datasets (DHS and MICS) limit the scope of the analysis to only those indicators for which data are collected. In reality, there are many variables shaping outcomes or access to opportunities. For example, the quality and reliability of a water connection is an important factor that might affect the access to basic drinking water. Similarly, distance from a health-care provider is an important circumstance that might shape women's access to skilled birth attendance during childbirth. These variables are not easily available in DHS and MICS surveys, so results have to be understood with this caveat.

Consistent with similar studies on inequalities among groups, this analysis also does not consider inequality within groups. Even with homogeneous groups (e.g. women from poorer households and with lower education), additional unobserved circumstances affect outcomes. This analysis only calculates the observable average outcome or access to an opportunity for each group, and thus draws conclusions on gaps and inequality based on these averages.

An important limitation is the lack of information on the income of individuals or households, as it is not collected by DHS and MICS. Instead, the analysis uses the wealth index, a composite index reflecting a household's cumulative living standard, developed by the DHS and MICS researchers. The wealth index combines a range of household circumstances including: a) ownership of household assets, such as TVs, radios and bicycles; b) materials used for housing; and c) type of water and sanitation facilities. The wealth index is calculated using the Principal Component Analysis and thus allows a relative ranking of households based on their assets.[4] The wealth index is not comparable across countries, as it may consist of different assets in each country. As a result, any cross-country comparison of households access based on "wealth" should be understood with that caution.

Finally, the results are limited by available indicators. The CART Analysis only presents circumstances in the tree branches if they are found to reduce "entropy". Ultimately, these circumstances define the composition of the groups, but should not be interpreted as "causes" of a lower access. There are also many other factors that could potentially impact the analysis, but because of the limitation of the datasets, have not been included.

3 Based on the CART methodology results, some examples of policy recommendations are presented in (ESCAP and ECLAC, 2019)2019.

4 Wealth Index Construction, https://www.dhsprogram.com/topics/wealth-index/Wealth-Index-Construction.cfm.

References

Kelleher, J. D., Mac Namee, B. and D'Arcy, A. (2015). *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. Second. Cambridge, MA: The MIT Press.

Paes de Barros, R. *et al.* (2009). *Measuring Inequality of Opportunities in Latin America and the Caribbean*. Washington, DC: The World Bank (WB). Available at: http://hdl.handle.net/10986/2580.

Ravallion, M. (2016). *The Economics of Poverty: History, Measurement, and Policy*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780190212766.001.0001.

Rawls, J. (1972). *A Theory of Justice*. Oxford: Clarendon.

Roemer, J. E. (1998). *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Sen, A. (1987). *Tanner Lectures in Human Values: The Standard of Living*. Cambridge: Cambridge University Press (Tanner Lectures in Human Values). doi: 10.1017/CBO9780511570742.

Shorrocks, A. F. (2013). 'Decomposition procedures for distributional analysis: a unified framework based on the Shapley value', *Journal of Economic Inequality*, 11(1), pp. 99–126. doi: 10.1007/s10888-011-9214-z.

UN Economic and Social Commission for Asia and the Pacific (ESCAP) (2017). *Sustainable Social Development in Asia and the Pacific: Towards a People-Centred Transformation*. Available at: https://www.unescap.org/publications/sustainable-social-development-asia-and-pacific-towards-people-centred-transformation.

__________ (ESCAP) (2018a). *Inequality in Asia and the Pacific in the era of the 2030 Agenda for Sustainable Development*. Available at: https://www.unescap.org/publications/inequality-asia-and-pacific-era-2030-agenda-sustainable-development#.

__________ (2018b). *Social Outlook for Asia and the Pacific: Poorly Protected*. Available at: https://www.unescap.org/publications/social-outlook-asia-and-pacific-2018.

__________ (2020). *Classification and Regression Trees (CART): A User Reference Guide for Identifying those Left Furthest Behind*. Available at: https://www.unescap.org/resources/leave-no-one-behind-decision-tree-user-reference-guide.

UN Economic and Social Commission for Asia and the Pacific (ESCAP) and UN Economic Commission for Latin America and the Caribbean (ECLAC) (2019). *Reducing inequality in FEALAC member countries: Innovative analysis to guide policymaking that leaves no one behind*. Available at: https://www.unescap.org/sites/default/files/FEALAC_Joint_Publication.pdf.

United Nations Sustainable Development Group (UNSDG) (2019). *Leaving No One Behind: A UNSDG Operational Guide for UN Country Teams (Interim Draft)*. Available at: https://unsdg.un.org/resources/leaving-no-one-behind-unsdg-operational-guide-un-country-teams-interim-draft.

United Nations (UN) (2015). *General Assembly Resolution 70/1, Transforming our world: The 2030 Agenda for Sustainable Development*. Available at: https://www.un.org/ga/search/view_doc.asp?symbol=A/RES/70/1&Lang=E.

__________ (UN) (2017a). *Leaving No One Behind: Equality and Non-Discrimination at the Heart of Sustainable Development*. New York. Available at: https://digitallibrary.un.org/record/1628748?ln=en.

__________ (2017b). *United Nations Development Assistance Framework Guidance*. Available at: https://unsdg.un.org/resources/united-nations-development-assistance-framework-guidance.

World Bank Group (WB) (2006). *Berlin Workshop Series 2006: Equity and Development*. Available at: http://hdl.handle.net/10986/6964.