

# THE LANCET

## Global Health

### Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Xu Y, Geldsetzer P, Manne-Goehler J, et al. The socioeconomic gradient of alcohol use: an analysis of nationally representative survey data from 55 low-income and middle-income countries. *Lancet Glob Health* 2022; **10**: e1268–80.

**Appendices**

**Contents**

**Text S1: Data search process ..... 2**

**Text S2: Search methods for low- and middle-income countries that did not have an eligible WHO STEPS survey..... 3**

**Text S3: Country-specific sampling methods ..... 4**

**Text S4: Detailed methodology for household wealth index calculation..... 29**

**Table S1: Population weighted descriptive statistics of participants in the final sample with complete data on current drinking (N=336,287) from nationally-representative population-based surveys conducted in 55 low- and middle-income countries between 2005 and 2017 ..... 31**

**Table S2: Multivariable logistic regression analysis investigating socio-economic gradient of current drinking and heavy episodic drinking stratified by sex ..... 32**

**Table S3: Association between missing information on heavy episodic drinking (HED) among current drinkers and individual-level socio-economic status, stratified by country income groups and sex, in 49 countries that queried information on HED..... 33**

**Table S4: A comparison of prevalence of current drinking by country: using information from the past 30 days and from the past 12 months..... 34**

**Table S5: The association between alcohol use and individual-level socio-economic status using multivariable logistic regressions adjusting for the interaction between country income levels and individual-level socio-economic status. .... 36**

**Figure S1: Predicted prevalence of heavy episodic drinking among male and female current drinkers at five levels of household wealth by country income group..... 38**

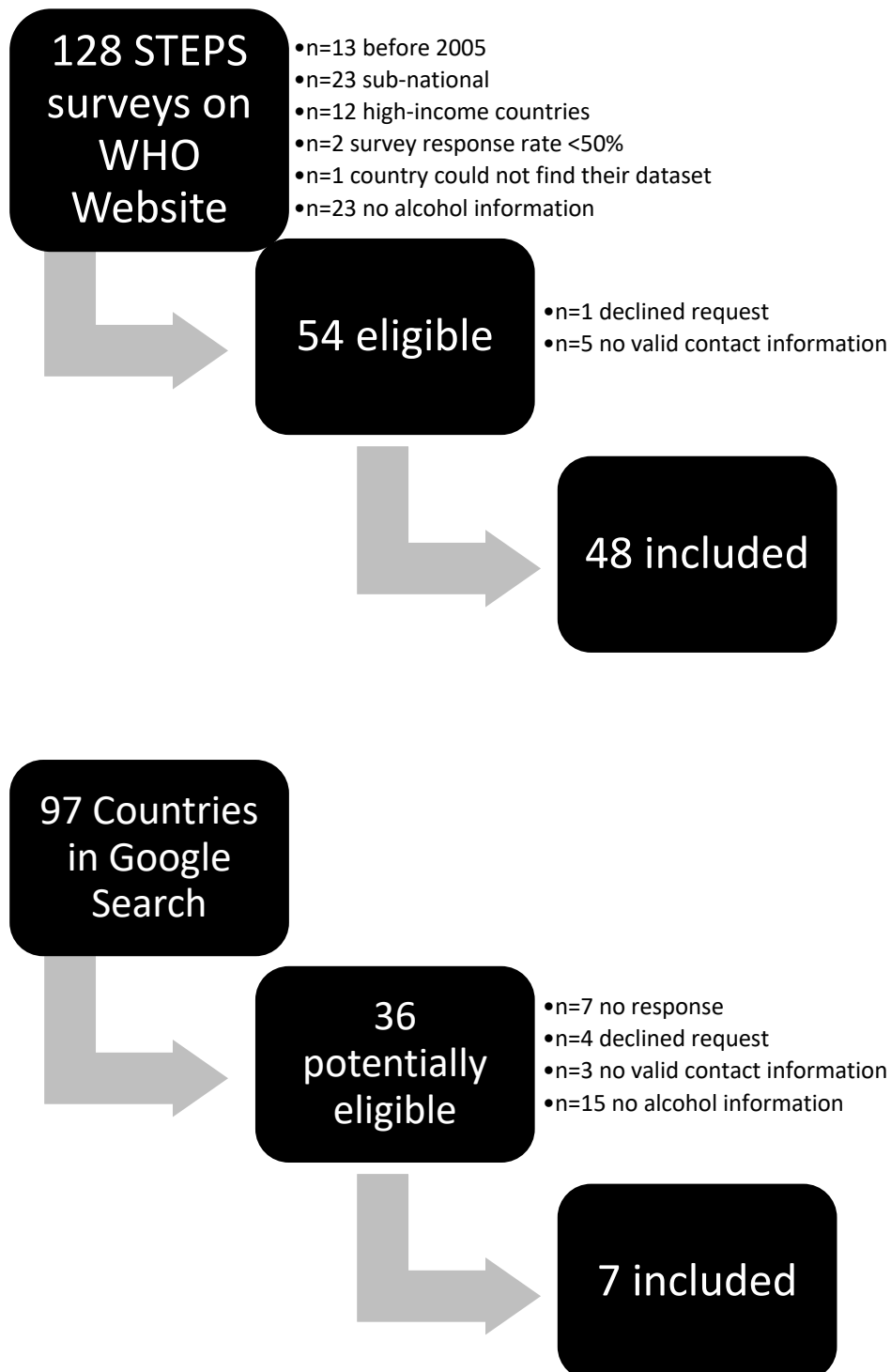
**Figure S2: Population-weighted prevalence of drinking patterns (daily drinking and heavy drinking) among drinkers in country income groups, stratified by sex ..... 39**

**Figure S3: Population-weighted prevalence of drinking patterns (daily drinking and heavy drinking) among drinkers in each group of individual-level socio-economic status (level of education and household wealth quintile), stratified by sex..... 40**

**Table S6: Multivariable logistic regression analysis investigating the association between heavy drinking and individual socio-economic status (level of education and wealth quintile), stratified by country income groups and sex. .... 41**

**Table S7: Multivariable logistic regression analysis investigating the association between daily drinking and individual socio-economic status (level of education and wealth quintile), stratified by country income groups and sex. .... 42**

## Text S1: Data search process



## **Text S2: Search methods for low- and middle-income countries that did not have an eligible WHO STEPS survey**

Search engine: Google

Search terms: “[country name]” AND (“population-based” OR household) AND (“alcohol” OR “drink”)

Number of hits reviewed: Hits reviewed until eligible survey identified, or, in the case of no eligible survey identified, first 500 hits (10 hits per page/5 pages reviewed)

### Inclusion criteria for a survey:

- (1) The survey was conducted during or after 2005; in cases where two surveys were available for a particular country, the most recent was used;
- (2) The survey data were made available at the individual level;
- (3) The survey contained alcohol use information in the past 12 months or in the past 30 days;
- (4) The survey was conducted in an upper-middle, lower-middle or low-income country according to the World Bank at the time the survey was conducted;
- (5) The survey was nationally representative;

Countries included in search: Afghanistan, Albania, American Samoa, Angola, Argentina, Armenia, Bangladesh, Belize, Bolivia, Bosnia and Herzegovina, Brazil, Bulgaria, Burundi, Cameroon, Central African Republic, Chad, Chile, China, Colombia, Congo, Cook Islands, Côte d'Ivoire, Cuba, Democratic People's Republic of Korea, Democratic Republic of the Congo, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Equatorial Guinea, Fiji, Gabon, Ghana, Guatemala, Guinea, Guinea-Bissau, Haiti, Honduras, India, Indonesia, Jamaica, Kazakhstan, Libya, Macedonia, Madagascar, Malaysia, Maldives, Mali, Marshall Islands, Mauritania, Mauritius, Mexico, Micronesia (Federated States of), Montenegro, Morocco, Namibia, Nicaragua, Nigeria, Occupied Palestinian Territory, Pakistan, Panama, Peru, Philippines, Romania, Russia, Saint Lucia, Senegal, Serbia, Somalia, South Africa, South Sudan, Suriname, Syrian Arab Republic, Thailand, Tunisia, Turkey, Turkmenistan, Ukraine, Venezuela, Yemen, Zimbabwe

### **Text S3: Country-specific sampling methods**

*Note: In order to ensure accuracy in reporting, sampling methods are pasted verbatim from specified sources.*

#### **Algeria STEPS 2016**

“A multi-stage cluster sample of households. One individual within the age range of the survey was selected per household.”

*Source: NCD Microdata Repository. Study Description• Available at: <https://extranet.who.int/ncdsmicrodata/index.php/catalog/91/study-description#page=sampling&tab=study-desc>*

#### **Belarus STEPS 2016**

“At the first stage of sampling, the sampling units in each stratum (urban / rural area) were enumeration plots formed for organizing the 2009 census. Within each region, strata (urban / rural area) from the total list of enumeration plots using systematic sample procedures with probability proportional to the size of the 2009 census plots, enumeration plots were selected. PSUs were selected independently for each stratum (urban / rural) within each region and city of Minsk using the following sample procedures: All census enumeration areas for the 2009 census of the Republic of Belarus were grouped within each region by urban and rural areas. Thus, a total of 13 counting plots were formed (2 groups in 6 regions and 1 group in Minsk). We will call each of these sets of enumeration plots a sampling segment. The enumeration plots within each sampling segment were ranked in the order of their geographical location, which was determined as a result of drawing a serpentine line on a geographical map. In order to cover the entire territory of the republic, a curve was drawn on the map, which in a serpentine order divided the territory of each region into layers, which were then assigned serial numbers depending on their geographical location. In all regions (except the city of Minsk), counting sections were assigned the number of the layer in which the corresponding settlement is located. In the city of Minsk, implicit stratification took place by means of a serpentine distribution directly of enumeration plots throughout the city, taking into account its administrative-territorial division. The ranking of the enumeration plots was carried out by increasing the number of the layer. For the size of the enumeration plot, information is drawn from to the 2009 census regarding the population in it (the number of members of private households, with the exception of HH located in closed areas). Using the ordered list of enumeration sites, the accumulated (cumulative) values of the enumeration area size (indicator “population”) were calculated. The final cumulative value represents the total population in the sample segment. (...) Using the "Random Number" function, a random value from 0 to 1 (Rh) is selected. The first selected enumeration area was determined by multiplying the selection step (Sh) by a random value (Rh) and comparing this value with the cumulative value of the “population” (Qh) indicator. Based on the selection step, all subsequent selected PSUs were determined. The selected counting sections i in the sample segment h were determined by comparing their cumulative values (Qhi) and the value (Uhi) (...)

The first stage of selection was completed by selecting the required number of PSUs in each of the seven regions, separately for urban and rural areas. In the republic as a whole, with different probabilities, 288 counting plots were selected: 144 in urban and 144 urban areas. (...) The selection of HHs was carried out within each selected counting area (PSU). The basis for the selection is a list of private households that include persons aged 18-69 years, indicating the addresses of residential apartments and households. (...) After that, from a list of HHs in each selected PSU, a fixed number of HHs was systematically selected, equal to 20, which corresponds to the established cluster size. The start of selection was determined randomly (using the "Random Number" function). Given the fact that the sizes of the counting sections are close, but all are different since the number of selected units is fixed, then in each PSU a new interval and a new random start of selection were calculated. In general, 5760 HHs were selected in the republic.”

*Source, translated from: Prevalence of Risk Factors of Non-Infectious Diseases in the Republic of Belarus STEPS 2016. Available at: <https://www.who.int/ncds/surveillance/steps/belarus/en/>*

**Belize: CAMDI 2005-2006**

“The CAMDI survey was a cross-sectional survey based on a probabilistic, stratified, multistage, cluster sampling design of the noninstitutionalized population of five Central American sites. The survey sampled included the entire national population in Belize; the overall metropolitan populations in San Jose, Costa Rica; Tegucigalpa, Honduras; and Managua, Nicaragua; and was restricted to the municipalities of Santa Tecla and Villa Nueva, which are part of the metropolitan areas of San Salvador and Guatemala City, respectively. In each city, the primary sampling unit was a cluster of independent households within predetermined geographic areas. The primary sampling units were grouped into geographic strata (sectors and compact segments or blocks). The sample was allocated proportionally to the size of the population within each geographic stratum of each city. All eligible individuals aged 20 years or older in the randomly selected households were invited to participate. Data were weighted to account for differential selection probabilities and survey nonresponse, and weights were poststratified to the adult population of each site based on age group and sex. The total sample population was 10,822, of whom 7,234 (66.8%) underwent anthropometry measurement and laboratory tests. Data were weighted to represent the population of the sampled city, except in Belize, where data were weighted to the country’s entire population.”

*Source: Barcelo A, Gregg EW, Gerzoff RB, et al. Prevalence of Diabetes and Intermediate Hyperglycemia Among Adults From the First Multinational Study of Noncommunicable Diseases in Six Central American Countries: The Central America Diabetes Initiative (CAMDI). Diabetes Care. 2012;35(4):738-740. doi:10.2337/dc11-1614.*

**Benin: STEPS 2015**

“The study was conducted using a three-stage random sampling technique. The sampling frame was provided by the National Institute of Statistics and Economic Analysis (INSAE) from the data of the fourth General Census of Population and Housing (RGPH4) in Benin in 2013. The first stage consisted of the random selection of 260 Enumeration Areas (EAs); The second stage consisted in drawing lots for 20 households per EA; The third step consisted in randomly selecting one individual per household retained according to the Kish method recommended by the WHO for the STEPS survey.”

*Source, translated from: Rapport final de l’enquête pour la surveillance des facteurs de risque des maladies non transmissibles par l’approche ‘STEPSSwise’ de l’OMSENQUETE ‘STEPS 2015’ au Bénin. Available at: <https://www.who.int/ncds/surveillance/steps/benin/en/>.*

**Bhutan: STEPS 2014**

“To achieve a nationally representative sample, a multistage sampling method was used to select enumeration areas, households and eligible participants at each of the selected households in three stages. The 2005 National Census was chosen as the basis for the sampling frame, with “Geogs” (blocks) in rural areas and towns in urban areas forming the primary sampling units (PSUs). Since the population distribution for urbanicity is 70:30 (rural:urban), 63 PSUs in rural and 14 PSUs in urban areas were chosen. PSUs were selected through the probability proportionate to size (PPS) sampling using the number of households in each PSU. Two secondary sampling units (SSUs) for every rural PSU and 4 SSUs for every urban PSU were selected. This led to the selection of 126 SSUs from rural and 56 SSUs from urban areas. This was also carried out by PPS sampling, using the number of households in each SSU. A total of 16 households from each SSU (both rural and urban) were selected using systematic random sampling. The sampling frame for this was the list of households with a unique identification number (ID) developed by the enumerators for the survey. At the household level, the Kish sampling method was used to randomly select one eligible member (aged 18–69 years) of the household for the survey. The Kish method ranks eligible household members in order of decreasing age, starting with males and then females, and randomly selects a respondent using the automated program for Kish selection in the handheld personal digital assistant (PDA).”

*Source: National survey for noncommunicable disease risk factors and mental health using approach WHO Steps Approach in Bhutan – 2014 Available at: <http://www.who.int/chp/steps/bhutan/en/>.*

#### **Botswana 2014**

“Botswana has a population of over 2 million with 27 districts and 4,845 enumeration areas and sample size of 300 enumeration areas with a target population of 6,400 people was systematically drawn from a pool of the whole enumeration areas. Against the identified enumeration areas numbers of households were listed and proportion of participants was calculated from the total sample size required for the country. Finally a computer generated random number was drawn to go into specific households in that specific enumeration area and at the end eligible participants residing in the household were listed into the electronic hand held data assistant(PDA) and at the end a name was picked automatically to participate in the survey.”

*Source: Botswana STEPS Survey Report on Non-communicable Disease Risk Factors - December 2015. Available at: <https://www.who.int/ncds/surveillance/steps/botswana/en/>.*

#### **Brazil: PNS 2013**

The Master Sample is a set of units of areas that are selected to meet various surveys of the IBGE Integrated System of Household Searches (SIPD). These units are considered primary sampling units (PSUs) in the sample planning of each of the surveys that use the Master Sample, such as PNS. The sampling plan consists of the stratification of the UPAs and selection of these units with probability proportional to the size, given by the number of permanent private households (DPPs).

The register for selection of the Master Sample was a file containing information from the Demographic Census 2010 on the census tracts of the geographic scope, whose limits are defined in the Operational Geographic Base 2010, totaling 316574 sectors. A sector or set of sectors with at least 60 DPPs was defined as UPA, with the exception of a few units, because it was not possible to aggregate sectors in some municipalities.

The stratification of the UPAs obeys four different criteria: administrative, including the division of the UF into capital, rest of the Metropolitan Region (RM) or Integrated Region of Economic Development - RIDE, and rest of the UF; geographical subdivision, which

subdivides capitals and other large municipalities into more strata; situation that involves rural / urban categorization; and the statistician in order to improve the accuracy of the estimates.

As part of the SIPD, the sampling design of the PNS followed, in part, the sampling design of the Master Sample, especially with regard to the stratification of the UPAs.

The PNS sample is by clusters in three stages of selection:

1st stage: selection with probability proportional to the size (given by the number of DPPs in each unit) of the UPAs sub-sample in each stratum of the Master Sample;

2nd stage: selection by simple random sample of households in each UPA selected in the first stage;

3rd stage: selection by simple random sampling of the adult (person aged 18 years or older) among all adult residents of the household.

The PNS will integrate the SIPD, which will make it possible to relate the information collected with other researches, such as the PNAD and the Household Budget Survey (POF) at different levels of geographic aggregation.

Source: translated from: Plano de Amostragem,

<https://www.pns.iciict.fiocruz.br/index.php?pag=planoamostragem>

### **Burkina Faso: STEPS 2013**

“Sampling methodology: The study was conducted on a sample obtained from a three-stage cluster stratified as recommended by the WHO for STEPS screening surveys. risk factors for noncommunicable diseases.

The sampling frame used was that derived from the general census of the population and habitat 2006 (RGPH 2006) and updated in 2010 during the survey Demographic and Health Survey of Burkina Faso (EDS-BF, 2010). This update concerned the enumeration areas (EAs) that correspond to the cluster as part of this study.



Selection of clusters: The choice of clusters was made according to a systematic random selection proportional to their size (in number of households) within strata (regions). To do this clusters were organized by stratum and place of residence (urban / rural). A total of 240 clusters of which 185 were in rural areas and 55 in urban areas were selected for the investigation.

Selection of households: Households were randomly drawn after an enumeration exhaustive list of all households in the cluster. A draw tool designed on Excel by the team. The technique was used in the field for selecting households to investigate. In total, 20 households in clusters were selected to participate in the study.

Selection of individuals: The choice of individuals was made randomly using Kish's method. In total, an individual aged 25 to 64 living in a selected household was fired for participate in the survey.”

*Source, translated from: Rapport de l'enquete nationale sur la prevalence des principaux facteurs de risques communs aux maladies non transmissibles au Burkina Faso Enquete STEPS 2013. Available at: [http://www.who.int/chp/steps/burkina\\_faso/en/](http://www.who.int/chp/steps/burkina_faso/en/).*

### **Cambodia 2010**

“The survey was designed to cover all geographical areas of Cambodia and a 3-stage sampling process as part of the multi-stage cluster sampling was carried out to randomly select the target population: random selection of communes (Khum in rural areas and its equivalent Sangkat in urban area) as primary sampling unit (PSU), followed by villages (Phum) for the secondary sampling unit (SSU), and by households for the elementary unit (EU). Finally all members of the randomly chosen households aged 25-64 years were invited to participate in this survey. The selection process was performed identically for urban and rural areas in order to get a self-weighted estimate for the whole population of the country. A total of 180 clusters with 34 clusters from the urban area and 146 clusters from the rural area were randomly selected.”

*Source: Prevalence of Non-Communicable Disease Risk Factors in Cambodia - STEPS survey Country Report, September 2010. Available at: <https://www.who.int/ncds/surveillance/steps/cambodia/en/>.*

### **Chile: NHS 2009-10**

“The sampling frame was constituted from the Population and Housing Census 2002. The design of the study was transversal, with a random sample of complex type households (stratified and multi-stage by clusters) with national, regional and area representation rural / urban. The target population was adults older than or equal to 15 years. The survey had a response rate in the eligible population of 85%. The refusal rate was of 12%. 5,434 people were interviewed. A nurse performed clinical and examinations to 5,043 participants and 4,956 accepted laboratory tests (blood and urine). The total sample loss of the oversized sample was 28% (this including rejection, non-contact and other causes of random loss). The raw sample was designed with overrepresentation of some population groups (older adults, regions other than the Metropolitan Region and rural areas) to increase sample efficiency and homogenize the accuracy of the estimators. The expansion of the sample data is because it grants each participant the weight that corresponds to it according to the design sample and at the same time corrects the distortion of the raw sample, making it coincide with the census population projection for January 2010 for Chilean adults over 15 years of age.”

*Source, translated from: Resumen Ejecutivo: Encuesta Nacional de Salud ENS Chile 2009-10. Available at: <http://epi.minsal.cl/encuesta-ens-antiores/>.*

### **Comoros: STEPS 2011**

“The STEPS survey on risk factors for chronic diseases in the Union of the Comoros took place from January to March 2011. This study has undertaken Step 1, Step 2 and Step 3.

Indeed, socio-demographic and behavioral measures were collected in Step 1. Physical measures such as height, weight and tension were collected in Step 2 and biochemical measurements were collected to assess the levels of blood glucose and cholesterol levels in Step 3. The STEPS survey conducted in Comoros Union is a survey of general population, targeting adults aged 25 to 64 years. A stratified survey was used to produce representative data for this age group. A total of 5556 adults aged 25 to 64 participated in the STEPS survey on a sample of 5760 people representing an overall response rate of 96.5%.”

*Source, translated from Union des Comores STEPS 2011 Note de synthèse.*

*Available at: <http://www.who.int/chp/steps/comoros/en/>.*

### **Costa Rica: STEPS 2010**

“The Costa Rican NCRFSS survey was a cross-sectional survey based on a probabilistic cluster sampling design. The NCRFSS survey was conducted during 2010 under the supervision of the Caja Costarricense de Seguro Social, a government public healthcare provider, and covers the overall adult population aged  $\geq 20$  years. Multistage cluster sampling was performed stratified by geographical areas, age groups (20–39, 40–64, and  $\geq 65$  years) and gender. The first sample stage was the randomized selection of the country’s geographical areas as primary sample units followed by the random selection of sectors in selected areas as secondary sample units. The random selection of areas and sectors was performed with probability proportional to size; the area or sector size was determined by the population  $>20$  years during 2009, as estimated by the Costa Rican Census and Statistics National Institute (INEC). Households were chosen through a random number generator using dwelling lists obtained from the health technician assistant in every community until all age group and gender strata sample sizes were achieved. A family dwelling was defined as a group of people who share the same table to eat. Survey participants were selected by the Kish method, which samples participants within a household with equal probability of selection, as recommended by the WHO STEPwise methodology. To be eligible for inclusion in the study, subjects had to be  $\geq 20$  years of age, permanently residing in the selected homes, and to have provided written consent. Pregnant or lactating mothers and those who were within 6 months postpartum were excluded from the study. Each participant selected for the study was informed of the study objectives and details before agreeing to participate in the investigation. In all, 3653 noninstitutionalized adults were surveyed, with an 87.8% response rate of the eligible population.”

*Source: Wong-McClure R, Gregg EW, Barcelo A, Sanabria-Lopez L, Lee K, Abarca-Gomez L, Cervantes-Loaiza M, Luman ET. Prevalence of diabetes and impaired fasting glucose in Costa Rica: Costa Rican National Cardiovascular Risk Factors Survey, 2010. J Diabetes. 2016 Sep;8(5):686-92.*

### **Ecuador: ENSANUT 2012**

Each province is divided into rural and urban strata and, additionally, two only urban strata were defined for Quito and Guayaquil. This results in 50 territories (dominio). As a first step, in each province 64 census blocks in rural and urban areas were chosen with probability proportional to size according to the number of households. In each block, 19 households were preselected of which 12 were finally selected. Finally, in each of the selected households, depending on the household composition, one woman in childbearing age was randomly selected as well as one person of each age group with adjustments according to the administration of the questionnaires. General household information and anthropometric measurements were taken from each household member. For a subsample of 50% biochemical measurements were taken and consumption information using a 24h recall diary was collected. The sample for the biochemical analysis consists of individuals aged 6 months to 59 years as well as pregnant women. However, there only were 231 pregnant women which is why they were dropped from the analysis of the report. The samples were collected by the Instituto Nacional de Estadística y Censos during 10 working days in different census zones. In total, samples of 21,520 individuals were collected which are 107.6% of the calculated subsample of 19,040 individuals.

Sources: Translated from pages 93 and 366 of MSP:ENSANUT-ECU\_06-10-2014

**Eritrea 2010**

“A multi-stage cluster sample of households. One individual within the age range of the survey was selected per household.”

Source: *NCD Microdata Repository. Study Description. Available at: <https://extranet.who.int/ncdsmicrodata/index.php/catalog/589/study-description#page=sampling&tab=study-desc>*

**Georgia: STEPS 2016**

“The STEPS survey of noncommunicable disease (NCD) risk factors in Georgia was carried out from June 2016 to September 2016. Georgia carried out Step 1, Step 2 and Step 3. Socio demographic and behavioural information was collected in Step 1. Physical measurements such as height, weight and blood pressure were collected in Step 2. Biochemical measurements were collected to assess blood glucose and cholesterol levels in Step 3. The survey was a population-based survey of adults aged 18-69. A Multi-stage cluster sampling design was used to produce representative data for that age range in Georgia. A total of 5554 adults participated in the survey. The overall response rate was 75.7%.”

Source: *Georgia STEPS Survey 2016 Fact Sheet. Available at: <http://www.who.int/chp/steps/georgia/en/>.*

**Ghana: SAGE 2007-8**

The sampling method used for the Ghana SAGE Wave 1 was based on the design for the World Health Survey, 2003, in which the primary sampling units (PSUs) were stratified by region and location (urban/rural). Selection of the PSUs was based on proportional allocation by size. Each enumeration area (EA) was selected independently within each stratum. In the WHS/SAGE Wave 0, a total of 6 000 households were to be interviewed and therefore 300 EAs were selected nationwide. Twenty households were to be randomly selected in each EA using systematic sampling. The number of EAs per region was based on the population size of the region. For SAGE Wave 1, a total of 5 000 50+ respondents and 1 000 18–49-year-old respondents were required and therefore 250 EAs out of the 298 EAs of the WHS/ SAGE Wave 0 were used based on the availability of respondents aged 50+ years within the EAs.

Enumeration areas with no 50+ individuals were not included. Within each EA, 20 households with one or more 50+ individuals and four households with members aged 18–49 were to be selected. All respondents aged 50+ within households with over 50s from the WHS

were automatically selected and additional households with members aged 50+ years were randomly selected to make a total of 20 households for each EA. The four households of the 18–49 years age group were randomly selected from the WHS/SAGE Wave 0 households list per EA. All the 50+ year olds within the selected households were to be interviewed together with the four identified under-50 respondents. Field work and data entry were undertaken between May 2007 and June 2008.

Source: Study on global AGEing and adult health (SAGE) Wave 1 The Ghana National Report Richard Biritwum, George Mensah, Alfred Yawson and Nadia Minicuci University of Ghana Medical School, Department of Community Health. July 2013

### **Grenada: STEPS 2010-11**

The sample frame comprised adults 25 to 64 years throughout Grenada, Carriacou and Petite Martinique. Administratively, the state of Grenada is divided into seven parishes with the islands of Carriacou and Petite Martinique being one parish. Each parish has a town (with the exception of St. David's) and several villages. A three-stage stratified sampling methodology was constructed using the Population and Housing Census 2001 as the sampling frame. The master frame was divided into 42 regions with an average size of eight (800) hundred households per region using a contiguous set of Enumeration Districts (EDs), where the approximate size of each ED is between 46 – 189 households (refer to Table 36 in the appendix). In the first stage, a paired design, i.e. samples of two EDs were randomly selected per region using a 3-digit table of random numbers. Since the frame was stratified by parish it would mean that in the large parishes, a larger number of EDs were selected. Since the ED size was fairly consistent, there was no need to use PPS sampling and hence simple random sampling was used. The number of households enumerated per ED was calculated

$((1736/42)/2) = 21$  There were therefore twenty-one households per selected ED. At the second stage, the sampling interval was computed by dividing the total number of households in the selected ED by the determined sample size of twenty-one (21) households per ED. Once the ED had been listed and the listing returned to the CSO a random table was used to select a random number (k) between 1 and the sample interval value, I, inclusive then to this number was added the sampling interval for the full list of households within the ED. Thus, the list of selected households was k, k+I, k+2I, ... k+(n-1)I where n is the size assigned to each ED (21). The third stage of the sampling required a listing of the members of the selected household then using the KISH Method the eligible person to be interviewed was selected.

#### **Guyana: STEPS 2016**

“A response rate of 66.68% will be selected based on the experience and response rates of other surveys over the years such as the recent Demographic Health Survey 2009. [...] STEPS 3 involve taking blood samples from a proportion of the sample, in this case 50% of the sample, in order to measure raised blood glucose levels and abnormal blood lipids. [...] The STEPS sample will be prepared by the Bureau of Statistics Guyana following the recommended STEPS sample methodology. A multi-stage cluster sampling design will be used. Guyana is divided into 10 administrative regions and within the administrative regions there are seven towns and each region is further divided into enumeration districts. For the STEPS survey 288 enumeration districts will be selected using the population probability sampling method and from each enumeration district 12 households will be selected giving a total sample size of 3456. Further at the household level each participant will be randomly selected by the electronic tablet. For STEP 3 50% of the sample will be randomly selected to participate. A re-listing of some households may also be necessary, such as those interior region locations, in which case in addition to household listings, enumeration districts maps will also be provided so that a re-listing can be done where required.”

*Source: STEPwise Approach to Chronic Disease risk factor surveillance (STEPS): Guyana's Implementation Plan. June 20, 2016. Ministry of Public Health, Guyana.*

#### **Iran 2016**

“For proportional to size sampling, we designed a systematic cluster random sampling frame through which 31,050 participants (3105 clusters) were selected from urban and rural areas of 31 provinces of Iran. To estimate the minimum sample at the 95% country with 384 samples (Ilam) was considered as the basis of calculations. The sample size of other provinces was calculated according to the population ratio of each to the referenced province. To consider the effect of sampling design and to control non-response error, 10% was added to the estimated samples of each province. With a view to reducing costs and increasing productivity, it was decided that for provinces with 800 or more samples through weighting methods, half of calculated sample size taken along with the twice weight in estimating. In this regard, national individual ID and postal code were used as part of individual characteristics in the questionnaire that has to be validated by interviewer through seeing national ID card.

The eligible population for study was defined according to the criteria of being among 18 years old Iranian adults that resided in Iran at the time of data collection. The first and second steps of study have been run for all selected samples and the third step was considered for those who were 25 years of age. Data were collected from individuals who agreed to participate and completed informed consent forms. The software features enabled us to analyze non-participation in each of study steps.”

*Source: Djalalinia, S et al., Protocol Design for Large-Scale Cross-Sectional Studies of Surveillance of Risk Factors of Non-Communicable Diseases in Iran: STEPs 2016. September 2017. Archives of Iranian Medicine, Volume 20, Number 9, pages 608-616*

### **Iraq 2015**

“A cross-sectional community based survey covering 15 governorates in Iraq. A Multi-stage cluster sampling technique was depended to select the minimum representative sample size to estimate the prevalence of the risk factors of noncommunicable disease through direct interview, physical examination and laboratory examination of blood samples of study participants. A total of 412 clusters were randomly selected each contain ten households. One subject from each household was randomly selected using KISH table to participate in the survey with a total sample size of 4120. Primary sampling units: The Sample was designed to provide estimates on a number of indicators on the situation of Noncommunicable diseases risk factors in Iraq at the national level. A national based rather than a governorate based sample is selected. A multi stage cluster sampling was used with stratification to urban and rural areas. Primary sampling units (PSUs) were the blocks, which consisted of 70 households or more before selection.”

*Source: Noncommunicable Diseases Risk Factors STEPS Survey Iraq 2015. Available at: <https://www.who.int/ncds/surveillance/steps/iraq/en/>.*

### **Kazakhstan: HHS 2012**

- KHHS 2012 uses a nationally representative sample consisting of 452 sample clusters and 13560 households, with 30 households to be interviewed per cluster;
- A cluster is a Census Control Area (CCA) created for the Population and Housing Census in 2009 (PHC 2009), with an average size of 1200 inhabitants per CCA;
- The sampling frame used for the KHHS 2012 is a complete list of all CCAs covering the entire country, provided by the Central Statistical Office of Kazakhstan;
- The target population of the survey: all residential households, all individuals living in residential households age of 15 years and over;
- The survey was expected to complete 12200 household interviews (assuming a household response rate 90%), and 11000 individual interviews of 15 years and over (assuming an individual response rate 90%); the survey interviewed only one eligible individual per household randomly selected from all the eligible;

- The survey is designed to provide representative results for all survey indicators for Kazakhstan; for the country's five geographical regions: West, North, East, South and Central; for the capital city Astana, and for the city of Almaty; for the urban and rural areas separately;
- The target sample was allocated to 16 Oblasts (Astana and Almaty are considered as special Oblasts) proportional to the Oblast population size with small adjustments.

Source: pasted verbatim from email exchange with study team.

### **Kenya: STEPS 2015**

“The 2015 Kenya STEPs survey was a national cross-sectional household survey designed to provide estimates for indicators on risk factors for non-communicable diseases for persons age 18 – 69 years. The sample was designed with a sample size of 6,000 individuals to allow national estimates by sex (male and female) and residence (urban and rural areas). The survey used the fifth National Sample Surveys and Evaluation Programme (NASSEP V) master sample frame that was developed and maintained by KNBS. The frame was developed using the Enumeration Areas (EAs) generated from the 2009 Kenya Population and Housing Census to form 5,360 clusters split into four equal sub-samples. A three-stage cluster sample design was adopted for the survey involving selection of clusters, households and eligible individuals. In the first stage, 200 clusters (100 urban and 100 rural) were selected from one sub-sample of NASSEP V frame. A uniform sample of 30 households from the listed households in each cluster was selected in the second stage of sampling. The last stage of sampling was done using Personal Digital Assistants (PDAs) at the time of survey, where one individual was randomly selected from all eligible listed household members using a programmed KISH method of sampling.”

*Source: WHO: Kenya STEPwise Survey for Non Communicable Diseases Risk Factors 2015 Report. Available at: [http://www.who.int/chp/steps/Kenya\\_2015\\_STEPS\\_Report.pdf?ua=1](http://www.who.int/chp/steps/Kenya_2015_STEPS_Report.pdf?ua=1).*

### **Kiribati 2015**

“As STEPS is intended to be nationally representative, a multi-stage cluster sampling method was used. The STEPS sampling spreadsheet was completed using the most recent census information (2012). The sample was selected in two stages assuming no replacement. At the first stage, a sample of Enumeration Areas (Islands and villages) from each stratum using probability proportional to size (PPS) sampling was selected. In the second stage, a fixed number of households from each selected Enumeration Area using systematic sampling was selected. The third stage of sampling selection was done at the household level using the KISH method. The sampling identified that data collection would be needed on the following islands: Makin, Butaritari, Marakei, Abaiang, North Tarawa, South Tarawa, Betio, Maiana, Abemama, Kuria, Aranuka, Nonouti, Tabiteuea North, Tabiteuea South, Arorae, Tabuaeran and Kiritimati.”

*Source: Kiribati NCD Risk Factors STEPS Report 2015-2016. Available at: <https://www.who.int/ncds/surveillance/steps/kiribati/en/>.*

### **Kyrgyzstan 2013**

“A multi-stage cluster sample of households. One individual within the age range of the survey was selected per household.”

Source: *NCD Microdata Repository. Study Description. Available at: <https://extranet.who.int/ncdsmicrodata/index.php/catalog/271/study-description#page=sampling&tab=study-desc>*

### **Lao People's Democratic Republic 2013**

“A multi-stage cluster sample of households. One individual within the age range of the survey was selected per household.”

Source: *NCD Microdata Repository. Study Description. Available at: <https://extranet.who.int/ncdsmicrodata/index.php/catalog/588/study-description#page=sampling&tab=study-desc>*

### **Lebanon 2017**

“A national cross-sectional survey adopting a two-stage cluster sampling design was conducted for Steps 1, 2 and 3. The sampling frames references used were the population distribution in Lebanon 2014, retrieved from the Central Administration for Statistics (CAS) and the Syrian population distribution data 2015, retrieved from UNHCR. 144 clusters were selected for the Lebanese sample and 144 clusters for the Syrian sample. The Primary Sampling Units (PSUs) were cadastral areas (cadasters) and the Secondary Sampling Units (SSUs) were the households. Twenty participants were recruited from each cluster. The latest available population estimates (cadastral data) were used, to randomly recruit PSUs by Probability Proportionate to Size (PPS). To account for the issue of the variability in the cadasters’ sizes, very small cadasters (<200 individuals) were combined with neighboring PSUs before selecting the sample, to enhance the likelihood of finding 20 target participants. On the other hand, cadasters with a large population size that were guaranteed to be sampled at least twice were handled as strata and each stratum were assigned a fixed number of random starting points based on how often it was selected with certainty. This was done using satellite images divided into grids, previously obtained from the Centers for Disease Control and Prevention (CDC)<sup>1</sup> for all Lebanese cadasters. For the Lebanese sample, the research team relied on the standard Expanded Program for Immunization (EPI) method for a systematic random selection of the households. Accordingly, within each selected PSU, households were identified using a systematic random approach following the WHO-UNICEF-EPI cluster method. The fieldworkers started with the highest floor on the right side of a building. If the household hosted an eligible participant, they proceeded with data collection, if not, they visited a second household which is selected by skipping 5 households. If during sampling, non-Lebanese households were selected, the fieldworker skipped them in a straight line until a Lebanese household was identified. This method has been previously used for national surveys in Lebanon. One participant was randomly selected within each household, using the eSTEPS application. Households were chosen until the target of 20 participants was reached. The PSUs for the Syrian refugees’ sample were identified, using the most recent available refugee estimates to randomly recruit PSUs by PPS. The same measures aforementioned were done to account for the variation in the cadasters’ sizes. The WHO-UNICEF- EPI cluster method was employed to select households. The fieldworkers targeted Syrian households; accordingly, when during sampling, non-Syrian households were selected, the fieldworker skipped them in a straight line until a Syrian household was identified. One participant was randomly selected within each household, using the eSTEPS application. For both samples, following STEPS’ team recommendations, sampling of participants was done without replacement, i.e. once a person was selected that person was not replaced with another one. Efforts were made to include all selected households. If the house was unoccupied at the time of the visit or if an adult was not available for an interview at the time of the visit, that house was revisited up to 4 times, with different visiting times. The number of refusals and nonresponses was recorded.”

Source: *WHO Stepwise Approach for Non-Communicable Diseases Risk Factor Surveillance Lebanon 2016-2017. Available at: <https://www.who.int/ncds/surveillance/steps/lebanon/en/>*

### **Liberia: STEPS 2011**

“Random multi-cluster sampling method was used to collect data during this survey in 5 of the 15 counties of Liberia with the district serving as the primary sampling unit. Different



sampling frames were designed and used at the district (Primary Sampling Unit-PSU), Chiefdoms (Secondary Sampling Unit-SSU) and household levels. Households listing generated from the 2008 National Population Census was used, and in each household, the list of individuals' resident was obtained and the Kish Method was used. Kish Method is a household sampling technique developed by WHO for STEPS. The field team selected households by using nutrition sampling method (throwing a pencil to get a selected direction). When the household enumeration sampling point is established, the interviewer counts all the households and using interval sample to get the household number. In each household, one person was selected using the Kish method."

*Source: WHO: The Final Report on the Liberia STEPS Survey 2011. Available at: [http://www.who.int/chp/steps/Liberia\\_2011\\_STEPS\\_Report.pdf?ua=1](http://www.who.int/chp/steps/Liberia_2011_STEPS_Report.pdf?ua=1).*

### **Moldova 2013**

"A two-stage cluster sampling procedure was carried out to select randomly participants from among the target population. Cluster sectors from the 2004 Moldova Population Census were used as a basic unit. Given the differences in lifestyle and disease status between populations in urban and rural areas, the target population was stratified into urban and rural areas of residence for the STEPS survey. At the first stage, within each stratum, primary sampling units (PSUs) (enumeration areas (EAs)) were selected systematically with probability proportional to the 2004 Population Census EAs (measure of size equal to the number of population in the EAs, provided by the census). Before selection, the census sectors were sorted geographically from north to south within each stratum, in order to ensure additional implicit stratification according to geographical criteria. A total of 400 clusters representing 400 EAs were selected from the 10 991 census EAs. These probabilistically selected clusters were used also in Moldova's DHS conducted in 2005, and the Multiple Indicator Cluster Surveys (MICS) conducted in 2012. Cartographic materials from the Population Census conducted in Moldova in 2004 were not available, thus it was not possible to use them for the STEPS survey. Therefore, for the first stage the probabilistic samples from the abovementioned surveys were used. Out of the 400 selected clusters, 167 were rural and 233 were urban. The distribution of the sample of 400 PSUs (EAs) for the DHS/MICS surveys was inversely proportional to the number of population within each stratum, taking into account that the response rate is lower in urban areas than rural owing to the smaller average size of the households in urban areas compared with rural areas. Thus, disproportional allocation with oversampling for urban areas was applied in the STEPS survey. A final weighting adjustment procedure was carried out to enable estimates at national and urban/rural levels. At the second stage, 15 households (secondary sampling units (SSUs)) were selected within each of the 400 PSUs. From the updated list of households used for the MICS 2012 survey, 15 households were selected randomly per cluster, using the Microsoft Excel® random sample tool. A total of 6000 individuals were selected from among the 400 clusters.

The Kish method was applied for the random selection of one individual aged 18–69 years from each household."

*Source: Prevalence of Noncommunicable Disease Risk Factors in the Republic of Moldova STEPS 2013. Available at: <https://www.who.int/ncds/surveillance/steps/moldova/en/>.*

### **Mongolia 2013**

"The survey was designed to cover all geographical areas of Mongolia, and a multi stage stratified sampling process was carried out to randomly select participants from the target population. Given the urban vs. rural differences in lifestyle and disease status, the target population was stratified into urban and rural areas and the sample was drawn proportionally based on the target population in each area. Ulaanbaatar, Darkhan and Erdenet cities represented urban areas, while the remaining aimags and soums represented rural areas. Primary units for Ulaanbaatar, Darkhan and Erdenet cities were khoros, whereas soums served as primary units for rural areas. The same principle used in the previous STEPS surveys in 2005 and 2009 was applied for sampling unit selections for each stage. From each selected household at the tertiary units of multi-stage cluster sampling in both urban and rural areas, only one individual aged 15-64 years old was randomly selected. The survey covered a total of 65 cluster sampling units. These units included randomly selected individuals from 32 soums in 21 rural aimags and 33 khoros in Ulaanbaatar, Darkhan and Erdenet cities. In order to be able to

compare the survey results and findings by urban and rural areas, we conducted sampling based on the principles to select approximately similar numbers of participants from both urban and rural areas.”  
*Source: Third national STEPS Survey on the Prevalence of Noncommunicable Disease and Injury Risk Factors-2013. Available at: <https://www.who.int/ncds/surveillance/steps/mongolia/en/>.*

#### **Morocco 2017**

“First stage (circulation of UP primary units): 244 (158 Urban Primary Units: UPU and 86 Rural Primary Units: UPR) primary units were drawn from the 4,500 UP constituting the master sample. Each UP is a geographic area with clear and identifiable boundaries on the ground and comprising on average 300 households. Second stage (drawing of secondary units (US) or clusters): When developing the master sample, the primary units were subjected to a cartographic division into zones of 50 households each. As a result, a UP has 6 secondary units on average. For the purpose of the STEPS survey, a single secondary unit or cluster was selected at the level of each sample UP. Each UP contains on average 6 clusters. Third stage (household circulation): Based on the cartographic sketch of the secondary unit (cluster) and starting from the starting point mentioned on the US limit sheet, and in a clockwise direction the interviewer draws the households to be surveyed by applying a step of two households. In each household, a random sampling of all individuals meeting the selection criteria was made, and a single participant is drawn at random in each household via the e-Steps application.”  
*Source, translated from: Enquête Nationale sur les Facteurs de Risque communs des Maladies Non Transmissibles 2017 – 2018 : Rapport. Available at: <https://www.who.int/ncds/surveillance/steps/morocco/en/>.*

#### **Mozambique: STEPS 2005**

“For the present community-based cross-sectional study, a sample of adults aged 25 to 64 years was assembled using the sampling frame of the 1997 census, which was designed to be representative at a national level and by place of residence (urban or rural).<sup>6</sup> Ninety-five geographical clusters were selected, among which all of the households were listed and 25 randomly selected and visited. All of the eligible subjects in the same household were invited for the study. Fifty-five subjects refused to participate, and 3323 were evaluated between September and November 2005.

Subjects were evaluated following the World Health Organization Stepwise Approach to Chronic Disease Risk Factor Surveillance (STEPS), which included a questionnaire on sociodemographic and behavioral factors (including smoking and drinking habits), medical and health history, and physical measurements (including blood pressure, weight, height, and waist circumference), using standardized methods.<sup>7</sup> The World Health Organization STEPS instrument for noncommunicable disease risk factors (core and expanded version 2.18) was used for data collection, after translation to Portuguese.

Blood pressure was measured on a single occasion by nonphysician trained interviewers using a semiautomatic sphygmomanometer (Omron 3) with an appropriate cuff size. After a 5-minute rest, blood pressure was measured twice, 1 minute apart, and a third measurement was performed if the difference between the first 2 was >10 mm Hg for systolic or diastolic blood pressure.

[...]

Anthropometric measurements were obtained with the participant wearing light clothing and no footwear. Body weight was measured to the nearest 0.1 kg using a digital scale and height to the nearest 0.1 cm in the standing position using a portable stadiometer.”

*Source: Albertino Damasceno, Ana Azevedo, Carla Silva-Matos, António Prista, Domingos Diogo, Nuno Lunet (2009). Hypertension Prevalence, Awareness, Treatment, and Control in Mozambique Urban/Rural Gap During Epidemiological Transition. Hypertension. 54: 77-83.*

#### **Myanmar 2014**

“To achieve a nationally representative sample, a multi-stage sampling method was used to select townships, wards and villages, households and eligible participants at each of the selected households. Stage 1: Selection of primary sampling units (PSUs) Administratively, Myanmar is divided into 330 townships. A township is subdivided into wards for urban settings and village tracts and then villages for rural settings. The list of townships has been used as the sampling frame at the first stage of sampling. Townships form the Primary Sampling Units (PSUs). Out of the total 330 PSUs, 52 PSUs were selected using Probability Proportionate to Size of population in each PSU (PPS). Stage 2: Selection of Secondary Sampling Units (SSUs) From each selected PSU (township), 6 SSUs (wards and villages) were chosen using probability proportionate to population size, totaling 312 SSUs for the whole country. Stage 3: Selection of eligible participants at household level From each selected SSU (ward/village), 30 households were selected using systematic random sampling. The sampling frame for this sampling is the list of households with unique identification number (ID) developed from a recent listing of households available from the Basic Health Staff. Stage 4: Selection of eligible participants at household level One eligible participant (aged between 25 and 64 years) in the selected households was recruited for the survey. The Kish sampling method was used to randomly select one eligible member of the household. Using the Kish Method, eligible participants (adults aged 25 to 64 years) in each household were ranked in order of decreasing age, starting with males then females, then randomly selected using the automated program for Kish selection in the handheld PDA. Each PSU (township) was estimated to contribute 180 participants, totaling 9,360 participants for 52 selected townships for the whole country. In actual study, the total sample size was 8757 participants.”

*Source: Report on National Survey of Diabetes Mellitus and Risk Factors of Noncommunicable Diseases in Myanmar (2014). Available at: <https://www.who.int/ncds/surveillance/steps/myanmar/en/>.*

### **Nepal: STEPS 2013**

“The surveyed population included men and women aged 15–69 years who had been living at their place of residence for at least six months. [...]

The sample size was calculated to represent the entire target population in Nepal. In order to achieve this statistical inference, the sample size calculator by WHO (sample\_size\_calculator STEPS) was used to derive a sample size of 4,200. [...]

Probability proportionate to size (PPS) was applied in the sampling strategy to improve the precision of the survey estimates. [...]

For this survey, the Ilaka was taken as the primary sampling unit (PSU). Out of the 921 Ilakas in Nepal, 159 are in the mountains, 467 in the hills and 295 in the Terai. The Steering Committee and the WHO NCD STEPS team at WHO headquarters in Geneva predetermined the number of PSUs to be taken in the study as 70. Thus, 70 Ilakas were sampled. Considering the varied distribution of the population across the ecological belts and to avoid the risk of under selection of the sample from the sparsely populated mountain belt, the distribution of Ilakas across ecological belts was determined on the basis of the population distribution pattern in the ecological belts (mountains 7%, hills 43% and Terai 50%). Hence, 30 Ilakas were selected from the hills, 5 from the mountains and 35 from the Terai using PPS. [...]

For the survey, wards (sub-units of VDCs and municipalities) were considered as clusters and taken as the secondary sampling unit (SSU). Three clusters were selected from each of the sampled Ilakas using PPS. All wards for each of the selected Ilakas were listed in order according to their numeric code, then 210 wards were selected (3 wards from each of the 70 Ilakas). To select the three wards from the list, all of the wards in the Ilaka were given a unique identification number, listed in ascending order along with household size and populated in the software. The software then selected the wards randomly on the basis of PPS.

Twenty households were selected from each cluster using systematic sampling. Thus, a total of 4,200 households were selected from the 210 clusters (20 households per cluster or ward). The sampling interval was determined by dividing the total number of households in the selected wards by 20. [...]

In municipalities, one ward covers a large number of households and each ward has more than 5 and sometimes up to 100 streets (margs or toles). Two margs or toles were selected and ten households were selected from each of the two margs or toles using systematic random sampling. If two or more families were found living in a house, one family was selected randomly. Eligible candidates (15–69 years) from the selected household were listed according to age and sex (males first and then females, in descending order), which was then fed into the Kish program in the personal digital assistants (PDAs), which automatically randomly selected one eligible candidate from each house.”

*Source: WHO: Non Communicable Diseases Risk Factors: STEPS Survey Nepal 2013. Available at: <http://www.who.int/chp/steps/nepal/en/>.*

### **Niger 2007**

“In the first stage, the EAs were taken from those delimited in the Third General Census of Population and Housing. The EAs were drawn up systematically and independently in each stratum. For each sample, the cumulative size of the population of each stratum was first calculated in the sampling frame. The Polling interval was then calculated as follows:  $I = M/a$ , rounded to the nearest whole, where  $M$  is the population of the stratum according to the sampling frame and the number of EAs to be drawn in the stratum. The series of survey numbers  $R, R + I, R + 2I, R + 3I$ , etc. was calculated, where  $R$  is the random number between 1 and  $I$ . Each survey number was then compared to the cumulative headcount column. The first EA that was drawn is the first EA on the list whose cumulative workforce is equal to or greater than the first survey number. The second EA that was drawn is the one that follows (after the first EA that has been drawn) whose cumulative number is equal to or greater than the second survey number, and so on. (...) 30 households were systematically drawn by EA. The step is therefore equal to the number of households in the EA divided by 304. In the

household, a person meeting the age criterion was interviewed, that is to say between the ages of 15 and 64. In the event that the quota for the 55-64 age group is difficult to reach, two participants have been selected in households where there is an adult aged 55 to 65, this in accordance with the Kish method.”

*Source, translated from: Mesure des facteurs de risque des maladies non transmissibles Au Niger (Approche Step"wise" de l'OMS). Available at: <https://www.who.int/ncds/surveillance/steps/niger/en/>*

## **Russian Federation: SAGE 2007-2008**

The SAGE Russian national sample was constructed using data from two sources:

1. The sample for the 2003 World Health Survey (WHS)
2. The 2002 All-Russia Population Census.

In constructing the SAGE Russian national sample, efforts were made to ensure even representation across administrative units. The largest administrative unit in the Russian Federation is the Federal District (FD). In 2007, there were seven FDs: Central, Southern, Northwestern, Urals, Volga, Siberian, and Far Eastern. Each FD is made up of federal subjects, administrative divisions which have varying levels of autonomy, but equal representation in the federal government; these include republics, krais, oblasts, federal cities, autonomous oblasts and autonomous okrugs. In 2007, there were 86 federal subjects. For the purpose of SAGE, federal subjects with particularly low population densities (making up 0.2% of the total population of Russia) were excluded: these included the Yamal-Nenets, Taimyr, Evenki, Koryak, and Chutkhotka autonomous okrugs, the republic of Sakha (Yakutia), Khabarovsk krai, and Magadan oblast.

### 2.1.1 First stage

The first stage of sample design was the definition of strata for selection. The sample was initially stratified by FDs, according to the distribution of population.

### 2.1.2 Second stage

The second stage of the sample design was selection of primary sampling units (PSUs), mainly according to the data from the 2002 Russian Census. For the selection of PSUs, first, all households which had participated in the WHS and which had a member aged 50-plus were listed again, as well as some WHS households with a member aged 18-49. These households were drawn from the three FDs that were included in the WHS: Central, Northwestern, and Volga.

Next, PSUs were selected for the remaining four FDs (Southern, Siberian, Urals and Far Eastern). Within FDs, administrative and territorial formations (ATF) constituted the primary sampling units (PSU). The population distribution within the four districts was used to determine the number of sample localities in each district, which were then weighted to reflect their representation in the four districts.

A computer program (according to the PPS method) was then used to select specific settlements in each district from the total number of ATFs, according to census data. ATFs were selected randomly and proportionally to the size of federal districts' population. This resulted in 39 ATFs from the Southern FD, 34 from the Siberian FD, 20 from the Urals FD and 7 from the Far Eastern FD. From each ATF, households, which constituted the study's secondary sampling units (SSU), were chosen at random, using a special formula for each territory sample. The probability of being including in the sample was equal for all households (0.00247). Between 1 and 551 households were selected from each ATF. Address lists for all selected households (including house and apartment numbers) were compiled with the help of out-patient clinic staff. In each territory, the sample was based on a household listing and enumeration, randomly selecting houses/ apartments until the desired sample size was reached. All members of each household selected for the survey sample were enumerated on the household roster and all eligible people aged 50-plus were invited to participate in the

survey. If a household had at least one person aged 50 or older, then that household was included in the 50-plus sample. In the remaining households (that is, with no member aged 50 or older) one respondent aged 18–49 was randomly selected using Kish tables (Kish, 1965; Kish, 1987).

In selected households, the individuals eligible for interview formed the ultimate sampling unit. The total sample size of individuals was targeted to be 1000 people in the age group 18–49 years old and 5000 people aged 50 or older.

### 2.1.3 Stratification and allocation of enumeration areas

From a total of 288 enumeration areas, 176 ATFs were visited: the coverage was 61.1% of the targeted ATFs, with the highest percent of visited enumeration areas in the Southern FD (71.8%). Of 7,200 eligible households, 4,644 HHs were included in the final sample and visited (both in urban and rural territories. 1,407 of them took part in the WHS survey.

Source: Study on global AGEing and adult health (SAGE) Wave 1 Russian Federation

National Report National Research Institute of Public Health, Russian Academy of Medical Sciences (RAMS). Study Report December 2013

### **Rwanda 2012**

“Multistage cluster sampling was used to select these participants from the population based on information from the last census. The three levels of clustering were: 1. Random selection of a statistical enumeration area (as defined by NISR) 2. Random selection of a household within the enumeration area 3. Random selection of an individual within the household. Selection of Enumeration Areas: Administratively, Rwanda is divided into thirty districts. In turn, each district is subdivided into sectors. Each sector is sub-divided into cells and then into villages. Villages are synonymous with enumeration area’s (EAs) in Rwanda and there are a total of 14,953 EAs in Rwanda. A total of 180 EA’s (or 1.2%) were randomly selected from this total using a probability proportional to size method that gives those EA’s with more people living in them a higher chance of being selected. In this way, the representativeness of the selected EAs is maximized. Selection of households: Forty households were randomly selected from within each of the selected EAs from a list of households supplied by NISR based on the most recent census. Selection of eligible participants: One eligible participant (an adult aged 15-64 years) was randomly selected from within each household using the Kish sampling method which is built into the PDAs used in the survey.”

*Source: Rwanda Non-communicable Diseases Risk Factors Report November 2015. Available at: <https://www.who.int/ncds/surveillance/steps/rwanda/en/>.*

### **Samoa 2013**

“The STEPS survey was a population-based survey of adults aged 18-64. A multi-stage, cluster sample design was used to produce representative data for that age range in Samoa. A total of 1766 adults participated in the survey. The overall response rate was 64%.”

*Source: Samoa STEPS Survey 2013 Fact Sheet. Available at: <https://extranet.who.int/ncdsmicrodata/index.php/catalog/707>*

### **Sao Tome and Principe 2008**

“The São Tomé and Príncipe STEPS survey is a survey of the general population, targeting adults aged 25 to 64. A cluster draw was used to produce representative data for this age group in São Tomé and Príncipe. A total of 2,457 adults participated in the São Tomé and Príncipe STEPS survey.”

*Source: São Tomé et Príncipe Enquête STEPS 2008 Note de synthèse. Available at: [https://www.who.int/ncds/surveillance/steps/sao\\_tome\\_and\\_principe/en/](https://www.who.int/ncds/surveillance/steps/sao_tome_and_principe/en/)*

### **Sierra Leone 2009**

“The multi-stage cluster sampling strategy was used in this study. The CSs as demarcated by Statistics Sierra Leone (SSL) were used as the first set of clusters i.e. primary sampling unit (PSU). Hundred CSs were selected using the probability proportionate to size (PPS) sampling method. The CSs contain several EAs which served as the secondary sampling units. Five hundred and fifty EAs were selected from within the selected CSs by the PPS sampling method. At the tertiary stage, at least ten households were selected from each selected EA by using a simple random technique. Finally, one eligible respondent was selected from the list all eligible respondents within a selected household using the Kish method as describe in the WHO stepwise approach to chronic diseases surveillance manual.”

*Source: The prevalence of the Common Risk Factors of Non-Communicable Diseases in Sierra Leone. Available at: [https://www.who.int/ncds/surveillance/steps/sierra\\_leone/en/](https://www.who.int/ncds/surveillance/steps/sierra_leone/en/).*

### **Solomon Islands 2015**

“A multi-stage cluster sample design was used to produce representative data.”

*Source: NCD Microdata Repository. Study Description. Available at: <https://extranet.who.int/ncdsmicrodata/index.php/catalog/710/study-description#page=sampling&tab=study-desc>*

### **Sri Lanka 2014**

“A multi stage cluster sampling method was used to select a nationally representative sample from the total population. Department of Census and Statistics of Sri Lanka performed the selection of the study sample. Population of each divisional secretariat (DS) divisions as per the preliminary results of the Census done in 2012 was used for sampling. Sri Lanka is administratively divided in to 9 provinces and 25 districts. Each district is divided to Divisional Secretariat (DS) areas. Each DS area is divided to many Census Blocks, and each Census Block consists of many households. Primary sampling unit (PSU):

The primary sampling unit (PSU) was a Divisional Secretariat (DS) area. Out of 331 DS areas available, 80 DS divisions were selected using proportionate to the size (PPS) sampling. Secondary sampling unit (SSU): A census block was considered as a SSU. From each DS division (PSU), six secondary sampling units (SSU) were selected using the proportionate to the size (PPS) sampling technique. Therefore, a total of 480 SSUs or census blocks were selected from 80 PSUs. Tertiary sampling unit (TSU): Number of houses in each census block depends on the area density and the population density in each DS division. Tertiary sampling unit (TSU) was the household and 15 households from each CB by random systematic sampling by the Department Census and Statistics. Therefore, a sample of 7200 (80x6x15) households were selected. In some instances, there were more than one household living in one house. People who are cooking and eating together were considered as one household. Whenever there were more than one household in a house, one



household was selected randomly to be included in the study. Selection of participants: Only one participant from each household was included in the survey. All the eligible members in the selected family were listed in descending order according to the age. Once this was done, these data was fed to the personal digital assistants (PDAs). The PDAs then automatically selected the eligible participant using the Kish method.”

*Source: Non Communicable Disease Risk Factor Survey Sri Lanka 2015. Available at: [https://www.who.int/ncds/surveillance/steps/sri\\_lanka/en/](https://www.who.int/ncds/surveillance/steps/sri_lanka/en/).*

### **St. Vincent & the Grenadines: STEPS 2013**

“The sample size was proportionately divided between the three main reporting strata (St.Vincent/Northern Grenadines/Southern Grenadines). The country’s most recent age breakdown is based on the 2001 national census, and was used to approximate the adult population 18-69 years by Island grouping. The survey was stratified by sex, age group (18-29, 30-44 and 45-69 years) and geographical location (St. Vincent/Northern Grenadines/Southern Grenadines). A three stage cluster sampling approach was used. Enumeration districts were randomly selected using Probability Proportional to Size (PPS) from the sampling frame. A total of 199 enumeration districts were selected. The sampling frame was developed using the number of households per enumeration district taken from the 2012 preliminary census report; enumeration districts had been subsequently revised (2010-2011) so that no enumeration district containing more than 150 Households would be randomly selected from the selected enumeration districts. Twenty-six (26) households per enumeration district were selected. Where an enumeration district had been split into 2 or more new enumeration districts the number of households in the previously defined enumeration district was divided equally between the newly revised enumeration districts. The household list for each selected enumeration district was updated prior to selection of households during a re-listing exercise.

Eligible persons at the household level were randomly selected using the Kish method. If no one is present in the selected household, a notification of visit card was left and the interviewer revisited at a later time. There were a total of three visits to a household before it was listed as non-response (i.e., one initial recruitment visit and two call backs). The person selected for interview must be at least 18 years on the last birthday but not older than 69 years old.

The collection of blood samples and the nutrition intake (24 hour recall) were also conducted at participants' homes. Collection of these data was completed during a morning revisit, during which participants were fasting.”

*Source: NHNS Implementation Plan July 10th 2013.*

### **Sudan 2016**

“A four-stage cluster sampling design was implemented. The four sampling stages were; 1) selection of states from the six regions 2) selection of clusters (a cluster was a Popular Administrative unit), 3) selection of households and 4) selection of eligible individuals. First Stage (State): Administratively Sudan is divided into 18 states which are grouped in six regions, (North, East, Khartoum, Central, Kordofan and Darfur region (Table 1). States were randomly selected from each region. No geographical areas or populations were excluded from the sampling frame. Thus 11 states were selected, probability proportional to the size, to represent the six regions. A list of the selected states is shown in Table 2.1. Second Stage (Cluster PAU): The Popular Administrative Units (PAU) is the smallest geographically border unit. These were defined as the ‘cluster’ in the region. Clusters were randomly sampled from all PAUs, from both urban and rural strata, according to probability proportional to size in each state, and urban/rural distribution. The PAUs inaccessible due to security conditions were not excluded from the sampling frame, because within certain areas the security status was continuously changing. However, it was planned that if a PAU was found to be inaccessible at survey time, it should be replaced by the nearest accessible unit. However, no replacement was

required during this survey. Third Stage (Household): Within the selected PAUs, all households (HH) were included in the sampling frame. Accordingly (HH) were selected using systematic random methods. Fourth Stage (Individual): The members of the household were first listed in the mobile application (customized software). The inclusion criteria for the listed members were: all individuals aged between 18 to 69 years, from both sexes, irrespective of his health status and living in the selected household for a minimum of 6 weeks. The application was then run and it randomly selected the individual who will be selected to participate in the study.”

*Source: Sudan STEPwise Survey for Non-Communicable Diseases Risk Factors 2016 Report. Available at: <https://www.who.int/ncds/surveillance/steps/sudan/en/>.*

### **Swaziland: STEPS 2014**

“A Multi-stage cluster sampling design was applied. The survey covered all the four regions of the country. The size of the country and the distances between the regions and communities made it possible for the survey to sample a population representing all the 4 regions. The Multi-stage sampling procedure was implemented in the following procedural steps:

Stage 1: All four regions were included as a sampling frame of our Primary Sampling Unit (PSU). The number of the PSUs at this stage ensured precision in the survey estimates and as a result 216 PSUs were selected using probability proportional to size sampling.

Stage 2: The second stage of cluster sampling procedure entailed listing, sorting and random systematic sampling of the Secondary Sampling Units (Households) within the PSUs selected in stage 1 where 20 households were selected from each PSU. Based on census data, only households with eligible participants were systematically sampled through random systematic sampling.

Stage 3: At this level, all the eligible participants within a household were sequentially listed into the PDAs and only one participant per household was randomly sampled using KISH method built into the PDAs. The KISH method is a widely used technique that uses a pre-assigned table of random numbers to identify the person to be interviewed.”

*Source: WHO STEPS: Noncommunicable Disease Risk Factor Surveillance Report Swaziland 2014. Available at: <http://www.who.int/chp/steps/swaziland/en/>.*

### **Tajikistan 2016**

“A multi-stage cluster sample of households. One individual within the age range of the survey was selected per household.”

*Source: NCD Microdata Repository. Study Description. Available at: <https://extranet.who.int/ncdsmicrodata/index.php/catalog/270/study-description#page=sampling&tab=study-desc>*

### **Tanzania: STEPS 2012**

"The STEPS survey in the United Republic of Tanzania was a population-based survey of adults aged 25-64. The study used both multistage cluster and random probability sampling procedures. Fifty of 119 total districts were randomly selected as primary sampling units (PSUs). Within these PSUs, enumeration areas (EAs) of > 50 households were randomly selected. Any EA with < 50 households was merged with a neighboring EA. Within the EAs, households were randomly selected from a list of all eligible households in the EA. A total of 5762 adults participated in the Tanzania STEPS survey. Within each selected household, the Kish method was used to select the STEPS participant. This procedure was followed until the predetermined sample was obtained for the enumeration area. The response rate for this survey was 94.7%.”

*Source: Tanzania STEPS Survey Report. Available at: [http://www.who.int/chp/steps/UR\\_Tanzania\\_2012\\_STEPS\\_Report.pdf?ua=1](http://www.who.int/chp/steps/UR_Tanzania_2012_STEPS_Report.pdf?ua=1)*

## Timor-Leste: STEPS 2014

“Note: Data from Census 2010 were used for all sampling considerations. Even though planning and mapping for 2015 Census is ongoing, data from the Census will only be available after July 2015.

### STEP 1: Selection of Enumeration Area

- (1) List of EA with number of HH by district for Census 2010 was obtained from the Directorate of Statistics. There are 1826 EAs in Timor-Leste. Out of these, 150 EAs were selected.
- (2) The number of EAs to be selected from each district was based on their proportion in the country's population as per Census 2010.
- (3) The numbers of Households (HH) per EAs varied from 0 to more than 300. Therefore, probability proportion to size (PPS) was used.
- (4) For each district, the EAs were arranged in ascending order of HH size.
- (5) Sampling interval was obtained by dividing the total number of HH in the district by the number of EA to be selected from that district.
- (6) A random number was generated between one and the sampling interval for that district, using tools available at random.org.
- (7) The EA where that random number fell was the first EA to be selected.
- (8) Subsequently, the sampling interval was added to the random number and the EA where this new number fell was selected. For the next number, the sampling interval was added to the number and so on, till the population of HH was exhausted or target number of EA achieved.
- (9) This was done separately for each district.
- (10) The final list was compiled and had 150 EAs. These are spread over about 125 sucos.

### STEP 2. Selection of Households in an Enumeration Area

#### Listing the house numbers to be visited

- (1) It was decided to use the 2010 HH size of each EA. Based on past experience, it was expected that the increase would be on an average about 4–5%.
- (2) The list of households to be selected by enumerators was decided centrally.
- (3) Sampling interval was calculated by dividing the total number of households in the EA by 18.
- (4) The first HH number was selected randomly by reading the last two digits of a currency note. If the number represented by the two digits was more than 18, the last digit was taken into consideration. For each EA, a different currency note was used. This could also be done it by using the tool at random.org. or by draw of lots.
- (5) The subsequent HH are identified by adding the sampling interval as was done for selection of EA.”

Source: *Timor-Leste STEPS Survey Report, [online]*  
at [http://www.who.int/entity/chp/steps/Timor-Leste\\_2014\\_STEPS\\_Report.pdf?ua=1](http://www.who.int/entity/chp/steps/Timor-Leste_2014_STEPS_Report.pdf?ua=1)

### **Togo: STEPS 2010**

“Those included in this survey are male or female subjects, living in urban or rural areas, aged 15 to 64 on the day of the survey, residing in the enumeration area for at least 6 months and having given their informed consent to participate in this study. [...] Three hundred clusters were randomly selected in a systematic draw with probability proportional to the size of the cluster (number of households) in the 4620 areas of enumeration of the DGSCN (General Directorate of Statistics and National Accounts) sampling frame. In order to obtain the 4,800 households at the rate of 1 individual / household, 16 households per cluster were randomly selected at the second stage of survey. In each of the selected households, one individual was selected as a survey participant via the Kish Method. A household was defined as the group of persons, who regularly share the main meal (regardless of their relationship). Households were not replaced in the event of a refusal or two unsuccessful visits to the eligible person selected by Kish's method. If the selected person was unwell or not present at the time of the interview, the investigators either tried to find a new appointment or searched for the respondent.”

*Source: Translated from WHO: The Final Report on the Togo STEPS Survey 2010. Available at: [http://www.who.int/chp/steps/2010STEPS\\_Report\\_Togo\\_FR.pdf?ua=1](http://www.who.int/chp/steps/2010STEPS_Report_Togo_FR.pdf?ua=1).*

### **Tonga 2017**

“The STEPS survey in Tonga was a population-based survey of adults aged 18-69 years. A Multi-stage clustering sample design was used to produce representative data for that age range. A total of 3858 adults participated in this Tonga STEPS survey. The overall response rate was 85.7%.”

*Source: Tonga STEPS Survey 2017 Fact Sheet. Available at: <https://extranet.who.int/ncdsmicrodata/index.php/catalog/713>*

### **Tuvalu 2015**

“The Tuvalu STEPS Survey was a population based cross-sectional survey of 18-69 year olds.”

*Source: NCD Microdata Repository. Study Description. Available at: <https://extranet.who.int/ncdsmicrodata/index.php/catalog/639/study-description#page=sampling&tab=study-desc>*

### **Viet Nam 2015**

“The sampling of STEPS was done in as part of the sampling for the (GATS) conducted in combination manner to save time and resources for these two surveys. Applied the multi-stages complex sampling process, the sampling process done by GSO was as follow: Sampling of clusters (EA) In the first stage of sampling, the primary sampling unit (PSU) was an enumeration area (EA). There are about 170,000 EAs in the whole Viet Nam and the average number of households in each EA is different between urban and rural areas. An average number of households in an urban EA and a rural EA is 133 households and 120 households, respectively. Sample of EAs were selected from the master sample frame. The master sample frame was a cluster frame made by the GSO based on the frame of Population and Housing Census 2009 and updated with data of 2014. Based on the Population and Housing Census data 2009, GSO prepared a 15% of master sample to serve as a national survey sampling frame. The master sample frame contains 25,500 enumeration areas (EAs) from 706/708 districts of Viet Nam (2 island districts were excluded from the GSO master sample frame). The master sample frame of GSO was divided by two stratification variables: urbanization (1 = urban; 2 = rural) and district group (1 = district/town/city of province; 2 = plain and coastal district; 3 = mountainous, island district). It means that the master sample frame was divided into 6 sample frames or 6 strata. The probability proportional to size (PPS) sampling method was used to select sample of EAs from 6 strata of master sample frame. The final sample of GATS included 315 EAs in the urban and 342 EAs for the rural. From these 657 EAs, 315 EAs were systematically selected for STEPS. Sampling of households: At the second stage of sampling, 10% households in each EA were selected. Thus, 15 households from the selected urban EA and 14 households from the selected rural EA were chosen using simple systematic random sampling. The total households for STEPS 2015 were 4,651 households. Sampling of individuals: One eligible person is then randomly selected from

each selected household for the STEPS 1 interview. The selection of individual is automatically done by the PDA program after eligible household members are entered into the PDA.”

*Source: National Survey on the Risk Factors of Non-Communicable Diseases (STEPS) Viet Nam 2015. Available at: [https://www.who.int/ncds/surveillance/steps/viet\\_nam/en/](https://www.who.int/ncds/surveillance/steps/viet_nam/en/).*

### **Zambia 2017**

“To ensure that the sample reflected the entire country of Zambia, a multi-stage cluster sampling technique was used to select a nationally representative sample of adults in Zambia aged 18 to 69 years. It was decided to utilize the household listing from the Zambia Population- Based HIV Impact Assessment (ZAMPHIA) - a household-based national survey that was conducted between March and August 2016 in order to measure the status of Zambia’s national HIV response. ZAMPHIA offered the most pragmatic up to date and accessible national household listing to be used as the sampling frame for this survey. The ZAMPHIA survey included 60,581 households drawn from 1,103 clusters referred to in this report as standard enumeration area (SEA). Thus the sample drawn for the STEPS survey was a subsample of the households selected for the ZAMPHIA survey. In the first stage of sampling, SEAs were selected from each province using probability proportional to size (PPS). In the second stage, 15 households in rural SEAs and 20 households in urban SEAs were selected systematically using appropriate sampling interval based on the number of households in that SEA. These households constituted the final list of households for the STEPS survey prepared for the field investigators (FI). In the third stage, while the FI approached the household and sought consent, all eligible members in the household were entered into the Android-based device used for the survey. The device then selected one member from the eligible members using a simple random sampling technique. The selected member was then interviewed having gone through the ethical process of consent after being provided with information on the survey. If the selected member was not available, a scheduled visit was made. If the selected member could not be reached after two scheduled visits he or she was considered as non-response. There was no replacement strategy so as to maintain the integrity and representativeness of the sample.”

*Source: Zambia STEPS for Non Communicable Diseases Risk Factors (2017). Available at: <https://www.who.int/ncds/surveillance/steps/zambia/en/>.*

### **Zanzibar: STEPS 2011**

“The sampling strategy used is a multi-stage cluster sampling with stratification. One hundred primary sampling units (PSU) were proportionately allocated proportionately across Zanzibar’s ten districts. Each district is divided into smaller geographical and administrative clusters (Shehia). These clusters are subsequently divided into smaller zones, which typically consist of 100-300 households. Zones smaller than 100 were merged to make a sufficiently large zone, and zones much larger than 300 were split into smaller zones. At the first stage of sampling, clusters were selected using simple random selection from the list of clusters within each district. At the second stage zones were randomly selected using probability proportionate to size. At the third stage, households were randomly selected from the household lists provided by the administrative leader of the Shehia. An additional seven households were randomly selected at the third stage, to account for ineligibility of certain households. The second and third stages of sampling were done using the WHO software STEPSsampling.xls. Finally, participants in eligible households were selected using the Kish method. A total of 2659 adults participated in the Zanzibar STEPS survey. The response rate was 97.6% of eligible participants. Of the included participants, the response rate for STEPS 3 was 98%.”

*Source: Zanzibar NCD Survey Report. Available at: [http://www.who.int/chp/steps/2011\\_Zanzibar\\_STEPS\\_Report.pdf](http://www.who.int/chp/steps/2011_Zanzibar_STEPS_Report.pdf).*

#### **Text S4: Detailed methodology for household wealth index calculation**

Countries using an asset index surveyed a range of assets, dwelling characteristics, and further country-specific variables. Utilizing the standard DHS approach, we used a principle component analysis to derive an asset index, from which we created unweighted wealth quintiles.

Countries using an income-based measurement mainly followed the STEPS template questionnaire put forward by the WHO. In this, respondents are being asked about the average earnings (taking the past year) of the household in a week, month, or year. In cases where this question is left unanswered, a pre-coded estimate of the households' annual income is to be indicated. This pre-coded estimate was usually expressed as quintiles and sometimes as categories that were defined by the countries' survey teams. Using both the pre-coded estimates as well as the continuous income, we again created unweighted wealth quintiles. In this, we assume that national incomes follow a log-normal distribution and make use of the procedure put forward by Harttgen and Vollmer (2013) in combining income quintiles and categories.

<b>Wealth Measure</b>	<b>Country</b>
Asset index	Brazil, Ecuador, Ghana, Kenya, Iran, Nepal, Russian Federation
Continuous income	Bhutan, Eritrea, Kazakhstan, Kiribati, Laos, Myanmar, Niger, Timor-Leste
Continuous income and quintiles	Algeria, Azerbaijan, Botswana, Cambodia, Comoros*, Georgia, Kyrgyzstan, Lesotho, Liberia*, Moldova, Rwanda*, Samoa*, Sao Tome and Principe*, Solomon Islands*, Swaziland*, Tanzania, Tajikistan, Togo*, Tuvalu, Zambia, Zanzibar
Continuous income and categories	Benin, Grenada, Guyana, Lebanon, Mongolia, Morocco
Income categories or quintiles only	

No income

Sri Lanka, Sudan, St. Vincent & the Grenadines,  
Tonga  
Belarus, Belize, Burkina Faso, Chile, Costa Rica,  
Iraq, Mozambique, Sierra Leone, Vietnam

---

\*Quintiles were not used as they displayed large discrepancies with respect to continuous income range or could not be correctly identified

Reference:

Harttgen, K., & Vollmer, S. (2013). Using an asset index to simulate household income.

Economics Letters, 121(2), 257-262.

**Table S1: Population weighted descriptive statistics of participants in the final sample with complete data on current drinking (N=336,287) from nationally-representative population-based surveys conducted in 55 low- and middle-income countries between 2005 and 2017**

Characteristic	Weighted Percent	Unweighted N
Educational Attainment		
No Formal Schooling	13•9 (11•9-16•1)	50,712
Less Than Primary School	16•1 (15•4-16•8)	47,045
Primary School Completed	14•0 (13•5-14•5)	57,528
Some High School	18•3 (17•3-19•3)	57,449
High School or Above	37•7 (35•4-40•1)	107,251
Missing		16,302
Household Wealth Quintile		
1 (Least Wealthy)	18•1 (17•2-18•8)	56,526
2	19•2 (18•4-19•9)	54,804
3	19•1 (18•1-20•1)	53,515
4	20•5 (10•6-21•4)	52,079
5 (Wealthiest)	23•2 (21•7-24•8)	50,219
Missing		69,144
Current Drinking		
No	63•4 (61•5-65•4)	229,323
Yes	36•6 (34•6-38•5)	106,964
Missing		0
Heavy Episodic Drinking		
No	54•0 (50•6-57•4)	40,575
Yes	46•0 (42•6-49•4)	23,346
Missing		272,366
Heavy Drinking		
No	90•6 (89•4-91•6)	52,951
Yes	9•4 (8•4-10•6)	4,946
Missing		278,390
Daily Drinking		
No	92•4 (91•5-93•2)	108,646
Yes	7•6 (6•8-8•5)	5,530
Missing		222,111



**Table S2: Multivariable logistic regression analysis investigating socio-economic gradient of current drinking and heavy episodic drinking stratified by sex**

	Odds Ratio (p-values)			
	Current Drinking		Heavy Episodic Drinking	
	Male	Female	Male	Female
<b>Educational Attainment</b>				
No Formal Schooling	0.76 (p=0.014)	0.53 (p<0.0001)	1.07 (p=0.633)	1.11 (p=0.661)
Less Than Primary School	0.89 (p=0.141)	0.47 (p<0.0001)	0.97 (p=0.796)	0.84 (p=0.368)
Primary School Completed	0.88 (p=0.030)	0.58 (p<0.0001)	0.85 (p=0.370)	0.99 (p=0.977)
Some High School	0.87 (p=0.060)	0.65 (p<0.0001)	0.90 (p=0.290)	1.05 (p=0.730)
High School or Above	Ref	Ref	Ref	Ref
<b>Household Wealth Quintile</b>				
1 (Least Wealthy)	0.83 (p=0.089)	0.54 (p<0.0001)	1.16 (p=0.691)	0.88 (p=0.677)
2	0.81 (p=0.004)	0.62 (p<0.0001)	1.03 (p=0.893)	0.72 (p=0.122)
3	0.92 (p=0.364)	0.66 (p<0.0001)	0.90 (p=0.667)	0.87 (p=0.197)
4	0.79 (p=0.021)	0.73 (p=0.001)	1.06 (p=0.735)	1.19 (p=0.316)
5 (Wealthiest)	Ref	Ref	Ref	Ref
Age, Age Square, Survey Year	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes

**Table S3: Association between missing information on heavy episodic drinking (HED) among current drinkers and individual-level socio-economic status, stratified by country income groups and sex, in 49 countries that queried information on HED.**

	Odds Ratio (p-values)					
	Low Income		Lower-middle Income		Upper-middle Income	
	Male	Female	Male	Female	Male	Female
<b>Educational Attainment</b>						
No Formal Schooling	1•02 (p=0•906)	0•51 (p<0•0001)	0•76 (p=0•084)	0•73 (p=0•062)	1•49 (p=0•234)	1•54 (p=0•522)
Less Than Primary School	0•86 (p=0•266)	0•63 (p<0•0001)	0•93 (p=0•635)	0•88 (p=0•254)	1•10 (p=0•600)	1•26 (p=0•301)
Primary School Completed	1•06 (p=0•687)	0•74 (p=0•020)	0•93 (p=0•455)	0•80 (p=0•255)	1•03 (p=0•863)	1•15 (p=0•386)
Some High School	1•24 (p=0•118)	0•83 (p=0•386)	0•85 (p=0•083)	1•18 (p=0•022)	1•14 (p=0•449)	1•00 (p=0•998)
High School or Above	Ref	Ref	Ref	Ref	Ref	Ref
<b>Household Wealth Quintile</b>						
1 (Least Wealthy)	1•23 (p=0•115)	0•81 (p=0•176)	0•87 (p=0•374)	1•28 (p=0•270)	1•55 (p=0•504)	0•13 (p=0•001)
2	1•19 (p=0•447)	1•09 (p=0•420)	0•86 (p=0•403)	1•46 (p=0•091)	1•24 (p=0•493)	0•19 (p=0•021)
3	1•28 (p=0•207)	1•03 (p=0•758)	0•99 (p=0•931)	1•71 (p=0•051)	1•05 (p=0•837)	0•17 (p=0•025)
4	1•18 (p=0•190)	1•33 (p=0•093)	0•78 (p<0•0001)	1•28 (p=0•043)	1•32 (p=0•532)	0•10 (p=0•003)
5 (Wealthiest)	Ref	Ref	Ref	Ref	Ref	Ref
Age, Age Square, Survey Year	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

**Table S4: A comparison of prevalence of current drinking by country: using information from the past 30 days and from the past 12 months.**

Country	Prevalence of Current Drinking (%)	
	Past 12 months	Past 30 days
Total	30.5 (28.7-32.3)	22.5 (21.2-23.9)
<i>Low Income Countries</i>		
Benin	37.3 (34.0-40.7)	27.4 (24.8-30.1)
Burkina Faso	31.3 (28.3-34.4)	27.9 (25.1-30.8)
Cambodia	64.0 (61.9-66.0)	54.1 (52.0-56.1)
Comoros	35.0 (23.9-47.9)	22.4 (13.0-35.7)
Eritrea	43.7 (39.8-47.8)	38.6 (34.9-42.5)
Liberia	29.0 (25.9-32.3)	23.3 (20.0-26.7)
Nepal	22.2 (19.8-24.9)	17.5 (15.3-20.0)
Niger	0.5	0.3
Rwanda	47.2 (44.9-49.6)	41.7 (39.4-43.9)
Sierra Leone	22.5 (18.5-27.2)	18.1 (14.8-21.9)
Tanzania	38.2 (35.5-40.8)	30.0 (27.5-32.6)
Togo	63.2 (60.5-65.8)	54.4 (51.8-57.0)
Zanzibar	3.4 (2.0-5.7)	1.8 (1.2-2.8)
<i>Lower-middle Income Countries</i>		
Bhutan	50.3 (47.2-53.3)	42.6 (39.5-45.7)
Georgia	69.5 (67.4-71.6)	39.3 (36.9-41.9)
Kenya	25.6 (22.7-28.8)	19.7 (17.2-22.5)
Kiribati	22.6 (14.6-33.3)	11.8 (8.3-16.5)
Kyrgyzstan	44.0 (38.6-49.5)	32.1 (26.8-37.9)
Laos	88.1 (84.6-90.9)	68.0 (63.6-72.2)
Lesotho	41.3 (38.5-44.1)	31.3 (28.3-34.4)
Moldova	79.2 (77.0-81.2)	62.2 (59.5-64.8)
Mongolia	64.9 (60.6-68.9)	36.8 (33.9-39.8)
Morocco	2.6 (2.1-3.4)	1.8 (1.2-2.4)
Myanmar	27.8 (24.2-31.8)	22.0 (19.1-25.3)
Samoa	18.4 (15.3-21.8)	11.1 (8.2-14.8)
Sao Tome and Principe	86.9 (83.7-89.6)	84.8 (81.0-87.9)
Solomon Islands	30.4 (27.5-33.5)	18.5 (15.8-21.6)
Sri Lanka	26.7 (24.9-28.6)	18.2 (16.8-19.7)
Swaziland	18.2 (16.0-20.6)	13.2 (11.5-15.1)
Tajikistan	6.7 (5.3-8.5)	5.1 (3.9-6.8)
Timor Leste	27.4 (25.0-29.9)	16.3 (14.5-18.2)
Vietnam	66.9 (64.9-68.8)	44.6 (42.3-46.9)
Zambia	27.4 (25.5-29.4)	22.3 (20.6-24.1)
<i>Upper-middle Income Countries</i>		
Algeria	3.4 (2.8-4.2)	2.2 (1.6-2.8)
Azerbaijan	21.5 (19.0-24.1)	14.1 (12.3-16.1)
Belarus	82.3 (80.0-84.4)	53.1 (50.4-55.7)
Botswana	34.7 (32.2-37.3)	26.7 (24.4-29.2)

Chile	100	79•0 (75•8-81•9)
Grenada	55	46
Guyana	57•3 (54•7-59•9)	41•2 (38•3-44•2)
Iran	4•2 (3•9-4•4)	2•1 (1•9-2•3)
Iraq	1•0 (0•5-1•6)	0•6 (0•3-1•2)
Kazakhstan	16•5 (15•7-17•2)	13•6 (12•9-14•4)
Lebanon	13•8 (10•6-17•8)	9•4 (6•9-12•6)
St• Vincent & the Grenadines	65•5 (59•5-71•0)	49•4 (45•2-53•6)
Tonga	14•7 (12•0-17•8)	6•5 (5•3-7•9)
Tuvalu	26•1 (18•9-34•8)	20•6 (14•3-28•6)

---

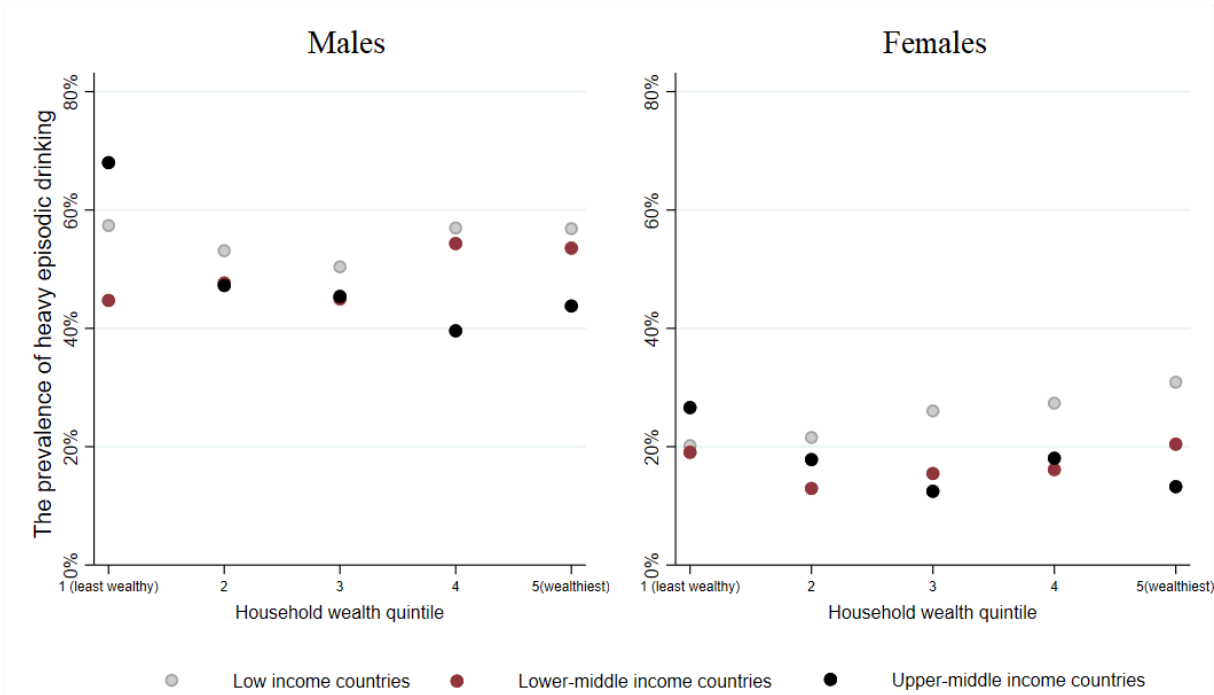
**Table S5: The association between alcohol use and individual-level socio-economic status using multivariable logistic regressions adjusting for the interaction between country income levels and individual-level socio-economic status.**

	Current Drinking		Heavy Episodic Drinking	
	Male	Female	Male	Female
<b>Educational Attainment</b>				
No formal schooling	0•40 (p<0•0001)	0•33 (p=0•001)	4•33 (p=0•066)	2•01 (p=0•294)
Less than primary school	0•77 (p=0•003)	0•38 (p<0•0001)	1•23 (p=0•399)	0•57 (p=0•694)
Primary school completed	0•91 (p=0•013)	0•59 (p<0•0001)	1•42 (p=0•332)	1•08 (p=0•114)
Some high school	0•83 (p=0•160)	0•67 (p<0•0001)	1•06 (p=0•863)	1•21 (p=0•158)
High school or above	Ref	Ref	Ref	Ref
<b>Education Attainment * Country GDP Class</b>				
No Formal Schooling * Low Income	2•07 (p=0•042)	2•14 (p=0•031)	0•24 (p=0•077)	0•67 (p=0•472)
No Formal Schooling * Lower-middle Income	2•13 (p=0•019)	1•87 (p=0•058)	0•22 (p=0•062)	0•55 (p=0•398)
Less than primary school * Low Income	1•33 (p=0•259)	2•09 (p<0•0001)	0•73 (p=0•298)	2•07 (p=0•776)
Less than primary school * Lower-middle Income	1•29 (p=0•202)	1•18 (p=0•461)	0•78 (p=0•367)	1•06 (p=0•723)
Primary school completed * Low Income	1•05 (p=0•844)	1•35 (p=0•018)	0•59 (p=0•160)	1•37 (p=0•377)
Primary school completed * Lower-middle Income	0•87 (p=0•288)	0•90 (p=0•489)	0•47 (p=0•045)	0•41 (p=0•322)
Some high school * Low Income	1•07 (p=0•769)	1•22 (p=0•054)	0•91 (p=0•786)	1•39 (p=0•250)
Some high school * Lower-middle Income	1•14 (p=0•645)	0•98 (p=0•817)	0•78 (p=0•508)	0•63 (p=0•466)
<b>Household Wealth Quintile</b>				
1 (least wealthy)	0•72 (p=0•011)	0•45 (p<0•0001)	3•28 (p<0•0001)	2•39 (p=0•317)
2	0•75 (p<0•0001)	0•61 (p<0•0001)	1•35 (p=0•453)	1•28 (p=0•130)
3	0•85 (p=0•195)	0•61 (p<0•0001)	1•27 (p=0•740)	0•91 (p=0•124)
4	0•70 (p<0•0001)	0•69 (p=0•004)	1•07 (p=0•890)	1•69 (p=0•153)
5 (wealthiest)	Ref	Ref	Ref	Ref
<b>Household wealth quintile * Country GDP Class</b>				
Lowest 20% * Low Income	1•31 (p=0•065)	1•48 (p<0•0001)	0•23 (p<0•0001)	0•28 (p=0•306)

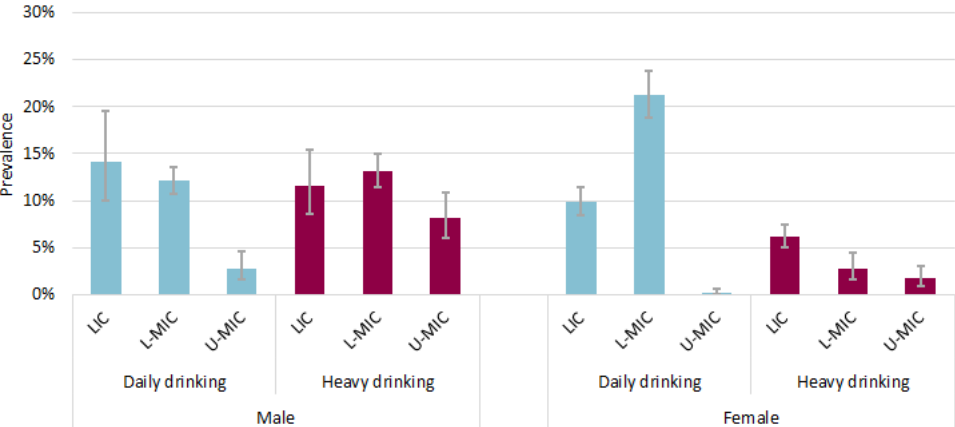
Lowest 20% * Lower-middle Income	1•40 (p=0•058)	1•61 (p=0•008)	0•20 (p=0•010)	0•34 (p=0•622)
20%-40% * Low Income	1•18 (p=0•231)	1•26 (p=0•053)	0•62 (p=0•252)	0•46 (p=0•144)
20%-40% * Lower-middle Income	1•18 (p=0•327)	0•87 (p=0•435)	0•64 (p=0•409)	0•23 (p=0•214)
40%-60% * Low Income	1•36 (p=0•058)	1•41 (p<0•0001)	0•56 (p=0•438)	0•94 (p=0•145)
40%-60% * Lower-middle Income	1•11 (p=0•506)	1•09 (p=0•522)	0•58 (p=0•464)	0•58 (p=0•524)
60%-80% * Low Income	1•44 (p=0•056)	1•30 (p=0•090)	0•84 (p=0•740)	0•66 (p=0•180)
60%-80% * Lower-middle Income	1•22 (p=0•229)	1•10 (p=0•628)	1•07 (p=0•892)	0•49 (p=0•482)
Age, Age Square, Survey Year	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes
Country GDP Class Fixed Effects	Yes	Yes	Yes	Yes

---

**Figure S1: Predicted prevalence of heavy episodic drinking among male and female current drinkers at five levels of household wealth by country income group**



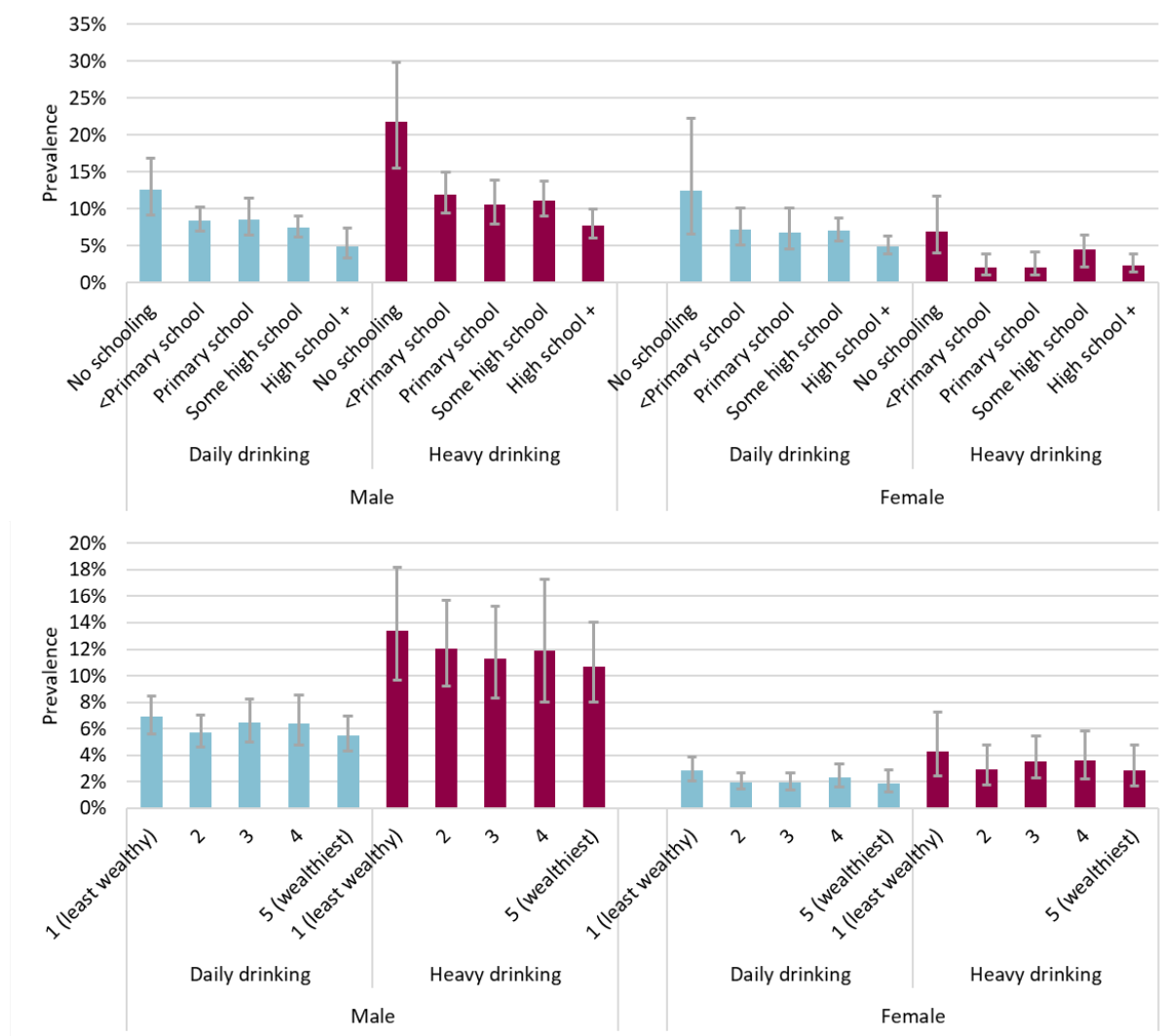
**Figure S2: Population-weighted prevalence of drinking patterns (daily drinking and heavy drinking) among drinkers in country income groups, stratified by sex**



*Note: LIC stands for low income countries, L-MIC stands for lower-middle income countries, and U-MIC stands for upper-middle income countries.*



**Figure S3: Population-weighted prevalence of drinking patterns (daily drinking and heavy drinking) among drinkers in each group of individual-level socio-economic status (level of education and household wealth quintile), stratified by sex**



**Table S6: Multivariable logistic regression analysis investigating the association between heavy drinking and individual socio-economic status (level of education and wealth quintile), stratified by country income groups and sex.**

	Odds Ratio (p-values)					
	Low Income		Lower-middle Income		Upper-middle Income	
	Male	Female	Male	Female	Male	Female
<b>Educational Attainment</b>						
No Formal Schooling	1•14 (p=0•575)	1•33 (p=0•512)	2•06 (p=0•015)	4•80 (p=0•064)	3•63 (p=0•148)	2•22 (p=0•001)
Less Than Primary School	1•17 (p=0•114)	1•08 (p=0•902)	1•33 (p=0•164)	1•26 (p=0•698)	1•93 (p=0•080)	0•38 (p=0•014)
Primary School Completed	1•24 (p=0•102)	1•19 (p=0•735)	1•05 (p=0•808)	2•62 (p=0•072)	0•89 (p=0•571)	1•48 (p=0•219)
Some High School	1•66 (p=0•048)	1•27 (p=0•534)	1•28 (p=0•044)	7•27 (p<0•0001)	1•14 (p=0•357)	0•96 (p=0•903)
High School or Above	Ref	Ref	Ref	Ref	Ref	Ref
<b>Household Wealth Quintile</b>						
1 (Least Wealthy)	1•18 (p=0•408)	1•11 (p=0•847)	1•63 (p=0•074)	2•42 (p=0•084)	10•57 (p=0•051)	1•23 (p=0•273)
2	1•08 (p=0•227)	1•24 (p=0•594)	1•75 (p<0•0001)	0•94 (p=0•803)	3•62 (p=0•050)	1•58 (p=0•155)
3	1•06 (p=0•692)	1•14 (p=0•707)	1•26 (p=0•042)	1•98 (p=0•072)	9•86 (p=0•044)	3•16 (p=0•999)
4	1•23 (p=0•386)	1•89 (p=0•184)	1•02 (p=0•931)	1•29 (p=0•464)	2•16 (p=0•077)	1•27 (p=0•143)
5 (Wealthiest)	Ref	Ref	Ref	Ref	Ref	Ref
Age, Age Square, Survey Year	Yes	Yes	Yes	Yes	Yes	Yes
Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

**Table S7: Multivariable logistic regression analysis investigating the association between daily drinking and individual socio-economic status (level of education and wealth quintile), stratified by country income groups and sex.**

		Odds Ratio (p-values)					
		Low Income		Lower-middle Income		Upper-middle Income	
		Male	Female	Male	Female	Male	Female
Educational Attainment							
	No Formal Schooling	1•64 (p=0•089)	4•59 (p<0•0001)	2•38 (p=0•001)	0•61 (p=0•239)	1•03 (p=0•969)	23•20 (p<0•0001)
	Less Than Primary School	1•44 (p=0•297)	3•33 (p=0•002)	1•74 (p=0•040)	1•13 (p<0•0001)	0•91 (p=0•776)	9•67 (p<0•0001)
	Primary School Completed	1•41 (p=0•292)	2•50 (p=0•078)	1•38 (p<0•0001)	1•37 (p<0•0001)	1•93 (p=0•027)	2•24 (p=0•340)
	Some High School	1•26 (p=0•015)	3•35 (p=0•002)	1•73 (p<0•0001)	1•59 (p<0•0001)	1•44 (p=0•444)	0•63 (p=0•387)
	High School or Above	Ref	Ref	Ref	Ref	Ref	Ref
Household Wealth Quintile							
	1 (Least Wealthy)	1•02 (p=0•946)	1•40 (p=0•492)	1•84 (p=0•007)	2•18 (p=0•077)	1•52 (p=0•330)	3•29 (p=0•048)
	2	0•86 (p=0•037)	1•10 (p=0•745)	1•49 (p=0•001)	1•59 (p=0•090)	1•28 (p=0•305)	4•83 (p=0•009)
	3	0•96 (p=0•723)	1•06 (p=0•878)	1•53 (p<0•0001)	1•24 (p=0•711)	1•70 (p=0•280)	0•37 (p=0•126)
	4	0•87 (p=0•010)	1•26 (p=0•460)	1•36 (p=0•045)	1•40 (p=0•421)	1•39 (p=0•285)	3•27 (p=0•269)
	5 (Wealthiest)	Ref	Ref	Ref	Ref	Ref	Ref
	Age, Age Square, Survey Year	Yes	Yes	Yes	Yes	Yes	Yes
	Country Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes