
Global School-Based Student Health Survey (GSHS)

User guide for R module for Sampling

Overview

- R code* has been developed to enable users to correctly draw a sample for their GSHS survey.
- The code ensures the sample will be representative and align with the sample design agreed upon with WHO.
- Upon completion of the sampling process, the user will receive a list of sampled schools as well as a school-level form for each selected school to perform the class-level selection.

**A Shiny app is available to enable users to use the R code with little to no knowledge of R.*

Background

- Prior to drawing a GSHS sample, the survey focal point must finalize the **sampling frame** (a standard template is available from WHO) and work with WHO to agree on a set of **sampling parameters**.
 - The sampling parameters define the desired number of schools and students to be sampled, any implicit or explicit stratification, and anticipated response rates at the school and student level.
 - Wherever implicit stratification is done, **sample allocation** calculations must be completed before proceeding with the sampling.
- Prior to 2021, all GSHS samples were drawn by US CDC using a purpose-built program called PC Sample (designed for Windows 95). The R code available from WHO performs the sampling just as it was done in PC Sample, thus ensuring continuity in GSHS methodology.

Preparation

- The sampling frame to be used by the R code must follow a very specific format. This is slightly different than the sampling frame template provided by WHO in that the headers and columns are somewhat simplified and standardized to enable the code to use the template easily.
- A detailed example of a sampling frame ready for the R sampling code is provided in the following slides.

Example Sample Frame File

- All sample frames must have the following fields (while Address is not mandatory, it's recommended to include Address as well as School ID or other identifying information to ensure the sample frame is clear and schools can be easily identified) :



school	address	A_BOYS	A_GIRLS	B_BOYS	B_GIRLS	C_BOYS	C_GIRLS	D_BOYS	D_GIRLS	E_BOYS	E_GIRLS	F_BOYS	F_GIRLS	Category
Red School	123 Red School St., Big City	14	6	15	4	16	9	27	7	20	3	9	5	national
Orange School	123 Orange School St., Medium City	0	67	0	50	0	69	0	0	0	0	0	0	national
Yellow School	123 Yellow School St., Big City	62	41	69	27	56	45	106	70	105	69	66	44	national
Green School	123 School St., Small City	22	29	11	24	7	21	9	19	0	18	0	15	national
Blue School	P.O. Box 123, Small Village	29	23	32	27	44	21	23	20	22	10	28	31	national
Purple School	123 Purple School St., Big City	41	21	28	26	33	19	0	0	0	0	0	0	national
Pink School	123 Pink School St., Medium City	0	105	0	101	0	83	0	110	0	68	0	55	national

NOTE! School names cannot include quotation marks.



In this example, category = national for all schools.
There is no implicit stratification done.

Example Sample Frame File



school	address	A_BOYS	A_GIRLS	B_BOYS	B_GIRLS	C_BOYS	C_GIRLS	D_BOYS	D_GIRLS	E_BOYS	E_GIRLS	F_BOYS	F_GIRLS	Category
Red School	123 Red School St., Big City	14	6	15	4	16	9	27	7	20	3	9	5	private
Orange School	123 Orange School St., Medium City	0	67	0	50	0	69	0	0	0	0	0	0	private
Yellow School	123 Yellow School St., Big City	62	41	69	27	56	45	106	70	105	69	66	44	private
Green School	123 School St., Small City	22	29	11	24	7	21	9	19	0	18	0	15	public
Blue School	P.O. Box 123, Small Village	29	23	32	27	44	21	23	20	22	10	28	31	private
Purple School	123 Purple School St., Big City	41	21	28	26	33	19	0	0	0	0	0	0	private
Pink School	123 Pink School St., Medium City	0	105	0	101	0	83	0	110	0	68	0	55	public

If IMPLICIT stratification is done, it would be reflected in the Category column. In this example, implicit stratification is by public vs private.

Example Sample Frame File

School

Address

Enrolment By Grade and Sex

Category

school	address	A_BOYS	A_GIRLS	B_BOYS	B_GIRLS	C_BOYS	C_GIRLS	D_BOYS	D_GIRLS	E_BOYS	E_GIRLS	F_BOYS	F_GIRLS	Category
Red School	123 Red School St., Big City	14	6	15	4	16	9	27	7	20	3	9	5	private_urban
Orange School	123 Orange School St., Medium City	0	67	0	50	0	69	0	0	0	0	0	0	private_urban
Yellow School	123 Yellow School St., Big City	62	41	69	27	56	45	106	70	105	69	66	44	private_urban
Green School	123 School St., Small City	22	29	11	24	7	21	9	19	0	18	0	15	public_urban
Blue School	P.O. Box 123, Small Village	29	23	32	27	44	21	23	20	22	10	28	31	private_rural
Purple School	123 Purple School St., Big City	41	21	28	26	33	19	0	0	0	0	0	0	private_urban
Pink School	123 Pink School St., Medium City	0	105	0	101	0	83	0	110	0	68	0	55	public_urban

Another example of a frame prepared for implicit stratification, this time by two characteristics: private vs public as well as urban vs rural.

Example Sample Frame File

The headers for the columns containing enrolment by grade and sex must be labeled as in the example below. The letter corresponds to the response option in the questionnaire. The letter must be followed by a double underscore and then “BOYS” or “GIRLS”. The number of columns will vary according to the number of grades/standards to be sampled in the survey.

3. In what grade are you?

- A. 7th grade
- B. 8th grade
- C. 9th grade
- D. 10th grade
- E. 11th grade
- F. 12th grade

School

Address

Enrolment By Grade and Sex

Category

school	address	A_BOYS	A_GIRLS	B_BOYS	B_GIRLS	C_BOYS	C_GIRLS	D_BOYS	D_GIRLS	E_BOYS	E_GIRLS	F_BOYS	F_GIRLS	Category
Red School	123 Red School St., Big City	14	6	15	4	16	9	27	7	20	3	9	5	national
Orange School	123 Orange School St., Medium City	0	67	0	50	0	69	0	0	0	0	0	0	national
Yellow School	123 Yellow School St., Big City	62	41	69	27	56	45	106	70	105	69	66	44	national
Green School	123 School St., Small City	22	29	11	24	7	21	9	19	0	18	0	15	national
Blue School	P.O. Box 123, Small Village	29	23	32	27	44	21	23	20	22	10	28	31	national
Purple School	123 Purple School St., Big City	41	21	28	26	33	19	0	0	0	0	0	0	national
Pink School	123 Pink School St., Medium City	0	105	0	101	0	83	0	110	0	68	0	55	national

Note: For grades with 0 enrolment in a school, it is ok to either put 0 or leave blank. The R code can handle either indication of zero enrolment for a given grade.

Sampling Parameters

- You will need to have your sampling parameters agreed with WHO before proceeding with the sampling. An example of sampling parameters is given below. The last three cells in green are demonstrative – these give an indication of the anticipated number of schools and students sampled after accounting for non-response. Your actual sample may vary from these figures.

General Information			Sample Parameters							
	Site Name	Field Year	Desired # of Schools	Expected School Response Rate	Desired # of Questionnaires	Expected Student Response Rate	Expected Overall Response Rate	# of Schools Sampled	# of Students Sampled	Average # of Students Sampled per Sampled School
GSHS	YourCountry	2024	25	0.8	2,500	0.8	0.64	32	3,125	125

- Accompanying these parameters would be any agreed stratification. It is usually also agreed to drop schools with less than 40 students.

Sampling Parameters (explicit stratification)

- In the case of explicit stratification, a sample would be drawn separately for each strata (i.e. you need to complete the entire sample selection process PER STRATA) and sampling parameters will be agreed per strata, like in the example below:

General Information			Sample Parameters							
Survey	Site Name	Field Year	Desired # of Schools	Expected School RR	Desired # of Questionnaires	Expected Student Attendance	Expected Overall RR	# of Schools Sampled	# of Students Sampled	Average # of Students Sampled per Sampled School
GSHS	YourCountry (Region A)	2024	25	0.8	2,500	0.8	0.64	32	3,125	125
GSHS	YourCountry (Region B)	2024	25	0.8	2,500	0.8	0.64	32	3,125	125
GSHS	YourCountry (Region C)	2024	25	0.8	2,500	0.8	0.64	32	3,125	125

Sample Allocation (for implicit stratification)

- Implicit stratification is often done to ensure the distribution of schools selected reflects the distribution of schools across the defined strata.
- The table below gives an example of how to allocate a targeted sample of 25 schools and 2500 students.
 - The enrolment figures are calculated from the sample frame. This is the total enrolment in the eligible grades across all schools in the strata.
 - The % column shows the % of students enrolled in each strata. These % values are then applied to the desired number of schools and students.
 - It is usually necessary to round the numbers of schools and students to be sampled, ensuring the targeted total is achieved.

	enrolment	%	# of schools to be sampled	schools (rounded)	# of students to be sampled	students (rounded)
Private	671,038	15%	3.872787	4	387.2787	387
Public	3,660,713	85%	21.12721	21	2112.721	2113
Total	4,331,751		25	25	2500	2500

Shiny app

- Once you have completed all preparations, you can proceed with the sampling in the Shiny app as follows:
 1. Complete ALL visible fields in the sampling app (see next slide)
 2. Upload the sampling frame
 3. Complete all fields that appear after sampling frame is uploaded (see slide after next)
 4. Click on the large button at the bottom of the app to proceed with the sampling and download the sampling output.

Shiny app

The screenshot shows the 'Sampling' section of a Shiny app. It includes a header 'Input site name and sampling parameters'. Below this, there are several input fields and sections:

- Site Name:** A text input field.
- School Response Rate:** A numeric input field with a value of 0.8. Callout: 'Anticipated school response rate'.
- Student Response Rate:** A numeric input field with a value of 0.8. Callout: 'Anticipated student response rate'.
- Parental Permission:** A numeric input field with a value of 1. Callout: 'Anticipated parental permission rate (leave as 1 if not requesting written permission)'.
- Minimum School Enrolment:** A numeric input field with a value of 40. Callout: 'Minimum school enrolment (usually use default of 40)'.
- Census?:** A radio button section with 'Yes' and 'No' options. Callout: 'Census?'.
- Double draw?:** A radio button section with 'Yes' and 'No' options. Callout: 'Double draw?'.
- Select language for school-level forms:** Radio buttons for 'English' (selected), 'French', 'Spanish', 'Russian', and 'Other'. Callout: 'Select desired language for school-level forms'.
- Upload sampling frame (NOTE: Must be in xlsx format):** A section with a 'Browse...' button and 'No file selected' text. Callout: 'Upload sampling frame'.
- Click here to download sampling output:** A button with a download icon. Callout: 'Proceed with sampling!'.

The World Health Organization logo is visible in the bottom left corner.

Shiny app

If your sample design **does not have implicit stratification**, you should see a single row of fields appear upon uploading your frame. The single Category value should reflect the single value of Category in your frame (e.g. “national”). Complete the number of questionnaires and number of schools to be sampled according to your sample parameters:

Upload sampling frame (NOTE: Must be in xlsx format)

Browse...

example_non-stratified_frame.xlsx

Stratified Sampling Inputs

Category	Number of Questionnaires Needed	Number of Schools to be Sampled
national		

If your sample design **does have implicit stratification**, you should see one row of fields per strata appear upon uploading your frame. The Category values should match the values of Category in your frame. Complete the number of questionnaires and number of schools to be sampled per strata.

Upload sampling frame (NOTE: Must be in xlsx format)

Browse...

example_stratified_frame.xlsx

Stratified Sampling Inputs

Category	Number of Questionnaires Needed	Number of Schools to be Sampled
JuniorHigh		
SeniorHigh		

Output

- Upon successful completion of the sampling, you will receive an Excel file containing the list of selected schools ...

Single Draw

school	enrolment	RevisedMOS	category	SchoolWeight	StudentWeight	Random Number	classes	School_Selected	Field_ID
A	307	307	national	5.763880771	3.023814572	0.1	3,6,9,12,15,1	Yes	1
B	247	247	national	7.164013752	2.43284104	0.125	3,5,7,9,11,13	Yes	2
C	209	209	national	8.466561707	2.058557803	0.05	1,3,5,7,9,11,1	Yes	3
D	182	182	national	9.722590092	1.792619714	0.055556	1,3,5,7,9,11,1	Yes	4
E	166	166	national	10.65970721	1.635026772	0.125	2,4,6,8,10,12	Yes	5
F	152	152	national	11.64152235	1.497132948	0.066667	1,2,3,4,5,6,7	Yes	6
G	134	134	national	13.20530893	1.319840888	0.15385	2,3,4,5,6,7,8	Yes	7
H	102	102	national	17.34815095	1.004655004	0.2	2,3,4,5,6,7,8	Yes	8

Double Draw

school	enrolment	RevisedMOS	category	SchoolWeight	StudentWeight	Random Number	All classes	School_Selected	Field_ID	GSHS	GYTS	AdjStudentWeight
A	342	342	national	2.637495831	1.982438015	0.029412	1,3,5,7,9,11,1	Yes	1	3,7,11,15,19,	1,5,9,13,17,2	3.96487603
B	307	307	national	2.938187538	1.779556932	0.033333	1,3,5,7,9,11,1	Yes	2	3,7,11,15,19,	1,5,9,13,17,2	3.559113863
C	261	261	national	3.45602902	1.512913222	0.038462	1,3,5,7,9,11,1	Yes	3	1,5,9,13,17,2	3,7,11,15,19,	3.025826444
D	247	247	national	3.651917304	1.431760789	0.16667	4,5,6,7,8,9,1	Yes	4	5,7,9,11,13,1	4,6,8,10,12,1	2.863521577
E	218	218	national	4.137722817	1.26365932	0.047619	1,2,3,4,5,6,7	Yes	5	2,4,6,8,10,12	1,3,5,7,9,11,1	2.527318639
F	209	209	national	4.315902269	1.211489898	0.15	3,4,5,6,7,8,9	Yes	6	3,5,7,9,11,13	4,6,8,10,12,1	2.422979796
G	185	185	national	4.875803104	1.072371441	0.055556	1,2,3,4,5,6,7	Yes	7	1,3,5,7,9,11,1	2,4,6,8,10,12	2.144742882
H	179	179	national	5.039237844	1.037591827	0.058824	1,2,3,4,5,6,7	Yes	8	1,3,5,7,9,11,1	2,4,6,8,10,12	2.075183653

Output

- ... as well as a school-level form for each school.

GSHS School-Level Form

Site : My Country

School: Red School

School ID: 1

School participation status: ☐ Participating

☐ Refusing

What grades/sections are taught in this school? _____

What grades/sections are to be surveyed in this school? _____

Obtain a list of classes for the grades/sections to be surveyed. The list of classes must be such that each student in the surveyed grades/sections must be represented on the list ONE and ONLY ONE TIME. Number the classes on the list beginning with the first eligible class listed.

What is the TOTAL number of eligible classes? _____

For each number below (these will be different for each school), select the corresponding class on your list. Keep selecting classes until you reach the END of your class list. If the first number listed below is greater than the number of eligible classes in your list OR if you run out of numbers before you reach the end of your list of classes, contact WHO.

3,8,13,18,23,28,33,38

For each selected class, enter the Class ID (the number above that you used to select the class), the teacher's name and the class name.

Class ID	Teacher Name	Class name	Class ID	Teacher Name	Class Name

Sampling Interval: 4.86962643553512
Random Number: 0.088235