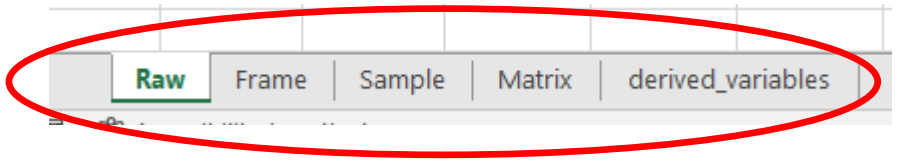
Global School-Based Student Health Survey (GSHS)

Data Processing and Weighting: Preparing the input file

World Health Organization

Overview

- In order to use the standardized R code for cleaning, weighting and analyzing your GSHS data, one Excel file with 5 sheets must be prepared:
    1. the raw dataset – named “Raw”
    2. the sampling frame – named “Frame”
    3. the sample – named “Sample”
    4. the questionnaire structure and desired indicators – named “Matrix”
    5. a description of any variables to be derived from multiple questions – named “derived_variables”
- Together these contain all the necessary information to perform the cleaning and weighting and produce the standard descriptive output.
- The naming, content and structure of these files must be as described in these slides. WHO colleagues can assist in constructing these files correctly and/or verify these files are correctly constructed prior to running the R scripts.



World Health Organization

Raw Dataset ("Raw" worksheet)

- The raw dataset file is a typical Excel dataset file with one row per student, one column per variable.
- Question responses should be A, B, C, etc. Do not recode these responses to numbers.
- ID variables for schools and classes must be named **school_id** and **class_id**
- Height must be in cms and weight must be in kgs. These variables must be named **height** and **weight**.

Raw Dataset ("Raw" worksheet)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	school_id	class_id	Page	q1	q2	q3	height	weight	q6	q7	q8	q9	q10	q11	q12	
2	1	4	2	E	A	B	166	55		C	B	C	B	B	B	C
3	1	4	3	E	A	B	165	50	C	C	B	G	F	B	A	E
4	1	4	4	F	B	B	158	60	C	G	B	A	D	D	A	E
5	1	4	5	F	A	B	167	56	C	C	C	D	B	B	B	C
6	1	4	6	D	A	B	161	46	C	C	B	B	B	B	C	E
7	1	4	7	E	A	B	168	60	C	B	D	D	D	D	G	C
8	1	4	8	F	A	B	169	60	C	B	B	C	E	B	B	C
9	1	4	9	F	B	B	164	72	C	D	D	E	B	C	A	D
10	1	4	10	F	B		152	55	B	A	A	C	B	A	B	C
11	1	4	11	F	B	B	153	55	C	C	G	B	B	B	B	D
12	1	4	12	F	B	B	156	50	A	B	D	D	E	A	C	D
13	1	4	13	E	A	B	170	56	E	B	B	C	E	D	B	E
14	1	4	14	D	A	B	160	54	C	G	E	D	D	D	A	C

World Health Organization

Sampling frame ("Frame" worksheet)

- The sampling frame is needed to complete the post-stratification adjustment of the analysis weights.
- It contains the number of students enrolled in each school in the original sampling frame by **grade** and **sex**.
- A **category** variable must be present indicating which stratum each school is in
  - If the sample was stratified, category = the name of the stratum, else all schools get the same value (e.g. "national")
- The number and ordering of grades **must match the question** students were asked about which grade they are in.
- At minimum, the name of each school must be included but it's advisable to include other identifying information such as address, school ID, etc.

World Health Organization

Note: The sampling frame spreadsheet used for drawing the sample can be reused here with no changes needed. This is by design.

Sampling frame ("Frame" worksheet)

3. In what grade are you?

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

Response options A-F appeared in the questionnaire.

The frame contains **2 columns per response** option containing the number of **boys** and the number of **girls** enrolled in that grade in that school.

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

The headers for the columns containing enrolment by grade and sex must be labeled as in the example above. 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.

World Health Organization

Sampling frame ("Frame" worksheet)

In this example, the sample was implicitly stratified by type of school. The **category** variable thus reflects this information.

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

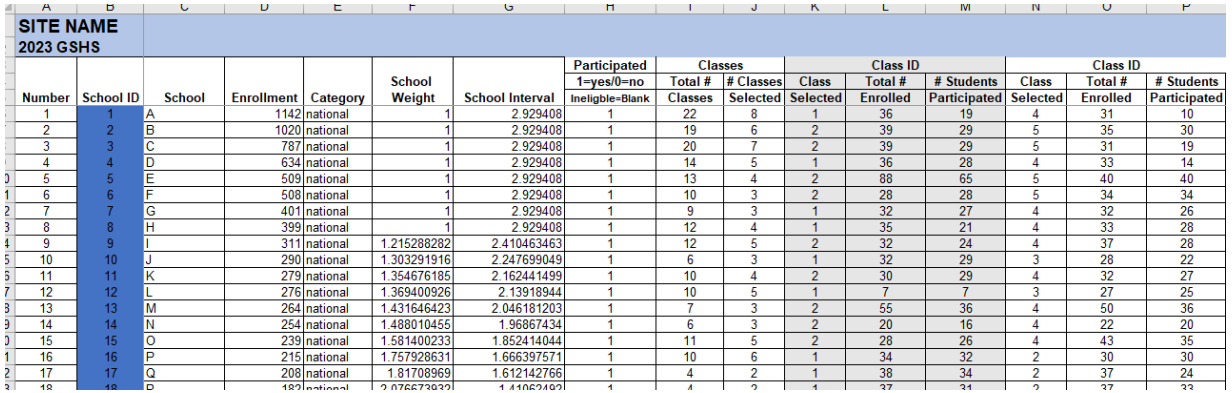
Important: Be sure all values of **category** are spelled correctly – misspellings would be interpreted as different strata (e.g. "private" would be interpreted as a different stratum than "privat")

World Health Organization

Sample (“Sample” worksheet)

- The sample file is used to calculate both the base weights and the non-response adjustments.
- It contains one row per **selected** school which is comprised of the variables in the table on the following slides.
- It is recommended to use the **login form template** available from WHO to log the responses from each school. Once completed, it will automatically create a worksheet matching the required structure of the “Sample” worksheet.

Login form template example:



SITE NAME 2023 GSHS															
							Participated	Classes		Class ID			Class ID		
							1=yes/0=no	Total #	# Classes	Class	Total #	# Students	Class	Total #	# Students
Number	School ID	School	Enrollment	Category	School Weight	School Interval	Ineligible=Blank	Classes	Selected	Selected	Enrolled	Participated	Selected	Enrolled	Participated
1	1	A	1142	national	1	2.929408	1	22	8	1	36	19	4	31	10
2	2	B	1020	national	1	2.929408	1	19	6	2	39	29	5	35	30
3	3	C	787	national	1	2.929408	1	20	7	2	39	29	5	31	19
4	4	D	634	national	1	2.929408	1	14	5	1	36	28	4	33	14
5	5	E	509	national	1	2.929408	1	13	4	2	88	65	5	40	40
6	6	F	508	national	1	2.929408	1	10	3	2	28	28	5	34	34
7	7	G	401	national	1	2.929408	1	9	3	1	32	27	4	32	26
8	8	H	399	national	1	2.929408	1	12	4	1	35	21	4	33	28
9	9	I	311	national	1.215288282	2.410463463	1	12	5	2	32	24	4	37	28
10	10	J	290	national	1.303291916	2.247699049	1	6	3	1	32	29	3	28	22
11	11	K	279	national	1.354676185	2.162441499	1	10	4	2	30	29	4	32	27
12	12	L	276	national	1.369400926	2.13918944	1	10	5	1	7	7	3	27	25
13	13	M	264	national	1.431646423	2.046181203	1	7	3	2	55	36	4	50	36
14	14	N	254	national	1.488010455	1.96867434	1	6	3	2	20	16	4	22	20
15	15	O	239	national	1.581400233	1.852414044	1	11	5	2	28	26	4	43	35
16	16	P	215	national	1.757928631	1.666397571	1	10	6	1	34	32	2	30	30
17	17	Q	208	national	1.81708969	1.612142766	1	4	2	1	38	34	2	37	24

World Health Organization

Sample ("Sample" worksheet)

Variable	Description	Required Name
School ID	Unique identifier of each school. It must be the same ID that is used in the raw data.	SCHOOL_ID
School weight	The inverse of the probability of selection of the school. This can be copied from the sample documentation.	SCWGT
Student weight (class sampling interval)	This is the interval used to derive the list of numbers on the school-level form. This can be copied from the sample documentation.	SCINTV
Total enrolment of the school for targeted grades	This can be calculated by summing all enrolment figures for the school in the sampling frame.	ENROLMENT
Sampling stratum of the school (category)	The name of the sampling stratum of the school. If no stratification was done, put the same value (e.g. national) for all schools.	CATEGORY

World Health Organization

Sample ("Sample" worksheet)

Variable	Description	Required Name
School participation flag	Set to 1 if school participated, otherwise set to 0.	SCHOOL_PART
Total number of eligible classes in the school	Refer to the completed school-level form for each school to obtain this number. This number will be written on the space intended for the TOTAL number of classes.	TOTCLASS
Total number of classes selected in the school	Refer to the completed school-level form for each school to obtain this number. This number can be determined by either looking to see how many numbers in the number list were circled or in the number of classes written in the small table just below the number sequence.	SELCLASS
For each selected class: class ID, total enrolment and number of participating students	Refer to the class-level form for each class for this information. Be sure to confirm that the class ID matches that used in the raw data AND the number of participating students matches the number of records for the class in the raw data.	CLASS1, CENROL1, STPART1 CLASS2, CENROL2, STPART2 etc.

World Health Organization

Sample (“Sample” worksheet)

In this example, 6 of the selected schools are shown. There are 5 sets of columns for classes as the number of classes selected in any of the schools is at most 5.

SCHOOL_ID	SCWGT	SCINTV	ENROLMENT	CATEGORY	SCHOOL_PART	TOTCLASS	SELCLASS	CLASS1	CENROL1	STPART1	CLASS2	CENROL2	STPART2	CLASS3	CENROL3	STPART3	CLASS4	CENROL4	STPART4	CLASS5	CENROL5	STPART5
1	273.3003054	11.68587179	1524	national	1	44	4	5	33	30	17	34	31	29	33	0	40	33	30	0	0	0
2	327.4447056	9.753562279	1272	national	1	36	4	7	36	34	16	36	36	26	35	34	36	32	31	0	0	0
3	364.0818754	8.772071735	1144	national	1	30	3	7	38	35	16	37	33	25	33	30	0	0	0	0	0	0
4	407.9428653	7.828920666	1021	national	1	30	4	3	36	36	10	35	35	18	36	34	26	36	36	0	0	0
5	483.1898671	6.609725381	862	national	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	648.7689493	4.922788509	642	national	1	24	5	3	36	33	8	32	29	13	35	21	18	26	25	23	28	28

World Health Organization

Sample (“Sample” worksheet)

One school did not participate, so its participation flag is set to 0 and the remaining columns in the row are all 0.

SCHOOL_ID	SCWGT	SCINTV	ENROLMENT	CATEGORY	SCHOOL_PART	TOTCLASS	SELCLASS	CLASS1	CENROL1	STPART1	CLASS2	CENROL2	STPART2	CLASS3	CENROL3	STPART3	CLASS4	CENROL4	STPART4	CLASS5	CENROL5	STPART5
1	273.3003054	11.68587179	1524	national	1	44	4	5	33	30	17	34	31	29	33	0	40	33	30	0	0	0
2	327.4447056	9.753562279	1272	national	1	36	4	7	36	34	16	36	36	26	35	34	36	32	31	0	0	0
3	364.0818754	8.772071735	1144	national	1	30	3	7	38	35	16	37	33	25	33	30	0	0	0	0	0	0
4	407.9428653	7.828920666	1021	national	1	30	4	3	36	36	10	35	35	18	36	34	26	36	36	0	0	0
5	483.1898671	6.609725381	862	national	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	648.7689493	4.922788509	642	national	1	24	5	3	36	33	8	32	29	13	35	21	18	26	25	23	28	28

World Health Organization

Sample (“Sample” worksheet)

One of the selected classes in school 1 was unable to participate, it is noted in the worksheet as having an enrolment of 33 but a participation of 0.

SCHOOL_ID	SCWGT	SCINTV	ENROLMENT	CATEGORY	SCHOOL_PART	TOTCLASS	SELCLASS	CLASS1	CENROL1	STPART1	CLASS2	CENROL2	STPART2	CLASS3	CENROL3	STPART3	CLASS4	CENROL4	STPART4	CLASS5	CENROL5	STPART5
1	273.3003054	11.68587179	1524	national	1	44	4	5	33	30	17	34	31	29	33	0	40	33	30	0	0	0
2	327.4447056	9.753562279	1272	national	1	36	4	7	36	34	16	36	36	26	35	34	36	32	31	0	0	0
3	364.0818754	8.772071735	1144	national	1	30	3	7	38	35	16	37	33	25	33	30	0	0	0	0	0	0
4	407.9428653	7.828920666	1021	national	1	30	4	3	36	36	10	35	35	18	36	34	26	36	36	0	0	0
5	483.1898671	6.609725381	862	national	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	648.7689493	4.922788509	642	national	1	24	5	3	36	33	8	32	29	13	35	21	18	26	25	23	28	28

World Health Organization

Analysis matrix (“Matrix” worksheet)

- The analysis matrix contains information about every question in the questionnaire, including:
  - The original question number used in the questionnaire (this needs to be formatted exactly as it is written in the raw dataset, e.g. “q1”)
  - The original question text and response options
  - The standard variable name assigned to this question
  - Information on the indicator(s) to be calculated from the question

- Each row of the matrix corresponds to one indicator, thus if multiple indicators are reported for a question, the question information is repeated.

Analysis matrix (“Matrix” worksheet)

The **site** column contains the question number from the fielded questionnaire. The values should match the variable labels in the raw dataset.

The **survey_question** and **var_levels** columns contain the standard variable name for the question, the question text and the response options.

	A	B	C	D	E	F	G	H	I
	bin_standard	site	numerator	denominator_resp_reduced	indicator_description	survey_question	var_levels	factsheet_section	factsheet_subtitle
2	DE_AGE	q1				DE_AGE: How old ar	A:11 years old or younger;B:12 years old;C:13 years old;D:14 ye		
3	DE_SEX	q2				DE_SEX: What is yo	A:Male;B:Female		
4	DE_GRADE	q3				DE_GRADE:In what	A:Class 7;B:Class 8;C:Class 9;D:Class 10;E:Class 11;F:Class 12		
5	DB_HEIGHT	height				DB_HEIGHT: How tall are you without your shoes on (in cm)?			
6	DB_WEIGHT	weight				DB_WEIGHT: How much do you weigh without your shoes on?			
7	DB_UNDERWT				DB_UNDERWT: Percentage of students who were underweight (<-2			dietary	Dietary Behaviours
8	DB_OVERWT				DB_OVERWT: Percentage of students who were overweight (>+1SI			dietary	Dietary Behaviours
9	DB_OBESE				DB_OBESE: Percentage of students who were obese (>+2SD from n			dietary	Dietary Behaviours
10	DB_B_FRUITNONE	q6	c('A')		DB_B_FRUITNONE: Percer	DB_FRUIT: During th	A:I did not eat fruit during the past 7 days;B:1 to 3 times during		
11	DB_B_FRUITLESS	q6	c('A','B','C')		DB_B_FRUITLESS: Percent	DB_FRUIT: During th	A:I did not eat fruit during the past 7 days;B:1 to 3 times during		
12	DB_B_FRUIT1	q6	c('D','E','F','G')		DB_B_FRUIT1: Percentage	DB_FRUIT: During th	A:I did not eat fruit during the past 7 days;B:1 to 3 times during		
13	DB_B_FRUIT2	q6	c('E','F','G')		DB_B_FRUIT2: Percentage	DB_FRUIT: During th	A:I did not eat fruit during the past 7 days;B:1 to 3 times during		
14	DB_B_FRUIT3	q6	c('F','G')		DB_B_FRUIT3: Percentage	DB_FRUIT: During th	A:I did not eat fruit during the past 7 days;B:1 to 3 times during		
15	DB_B_VEGNONE	q7	c('A')		DB_B_VEGNONE: Percent	DB_VEG: During the	A:I did not eat vegetables during the past 7 days;B:1 to 3 times		
16	DB_B_VEGLESS	q7	c('A','B','C')		DB_B_VEGLESS : Percenta	DB_VEG: During the	A:I did not eat vegetables during the past 7 days;B:1 to 3 times		
17	DB_B_VEG1	q7	c('D','E','F','G')		DB_B_VEG1: Percentage c	DB_VEG: During the	A:I did not eat vegetables during the past 7 days;B:1 to 3 times		

World Health Organization

Analysis matrix (“Matrix” worksheet)

The **bin_standard** column contain the name of the indicator.

The **indicator_description** column contains the name and text of the indicator derived from the question.

A	B	C	D	E	F	G	H	I
bin_standard	**site**	**numerator**	**denominator_resp_reduced**	**indicator_description**	**survey_question**	**var_levels**	**factsheet_section**	**factsheet_subtitle**
TO_B_DAYSCIG	q37	c('B','C','D','E','F','G')		TO_B_DAYSCIG: Percentage of students who currently smoked cigarettes (on at least 1 day during the 30 days before the survey)	TO_DAYSCIG: During the past 30 days, on how many days did you smoke cigarettes?	A:0 days;B:1 or 2 days;C:3 to 5 days;D:6 to 9 days;E:10 to 19 days;F:20 to 29 days;G:All 30 days	tobacco	Tobacco Use
TO_B_STOPCIG	q38	c('B')	c('A')	TO_B_STOPCIG: Percentage of students who tried to stop smoking cigarettes (among students who smoked cigarettes during the 12 months before the survey)	TO_STOPCIG: During the past 12 months, did you try to stop smoking cigarettes?	A:I did not smoke cigarettes during the past 12 months;B:Yes;C:No		

World Health Organization

Analysis matrix (“Matrix” worksheet)

The **numerator** column contains a list of response options which comprise the numerator of the indicator.

The **denominator_resp_reduced** column contains a list of response options to be excluded from the denominator.

	A	B	C	D	E	F	G	H	I
	bin_standard	**site**	**numerator**	**denominator_resp_reduced**	**indicator_description**	**survey_question**	**var_levels**	**factsheet_section**	**factsheet_subtitle**
	TO_B_DAYSCIG	q37	c('B','C','D','E','F','G')		TO_B_DAYSCIG: Percentage of students who currently smoked cigarettes (on at least 1 day during the 30 days before the survey)	TO_DAYSCIG: During the past 30 days, on how many days did you smoke cigarettes?	A:0 days;B:1 or 2 days;C:3 to 5 days;D:6 to 9 days;E:10 to 19 days;F:20 to 29 days;G:All 30 days	tobacco	Tobacco Use
	TO_B_STOPCIG	q38	c('B')	c('A')	TO_B_STOPCIG: Percentage of students who tried to stop smoking cigarettes (among students who smoked cigarettes during the 12 months before the survey)	TO_STOPCIG: During the past 12 months, did you try to stop smoking cigarettes?	A:I did not smoke cigarettes during the past 12 months;B:Yes;C:No		

World Health Organization

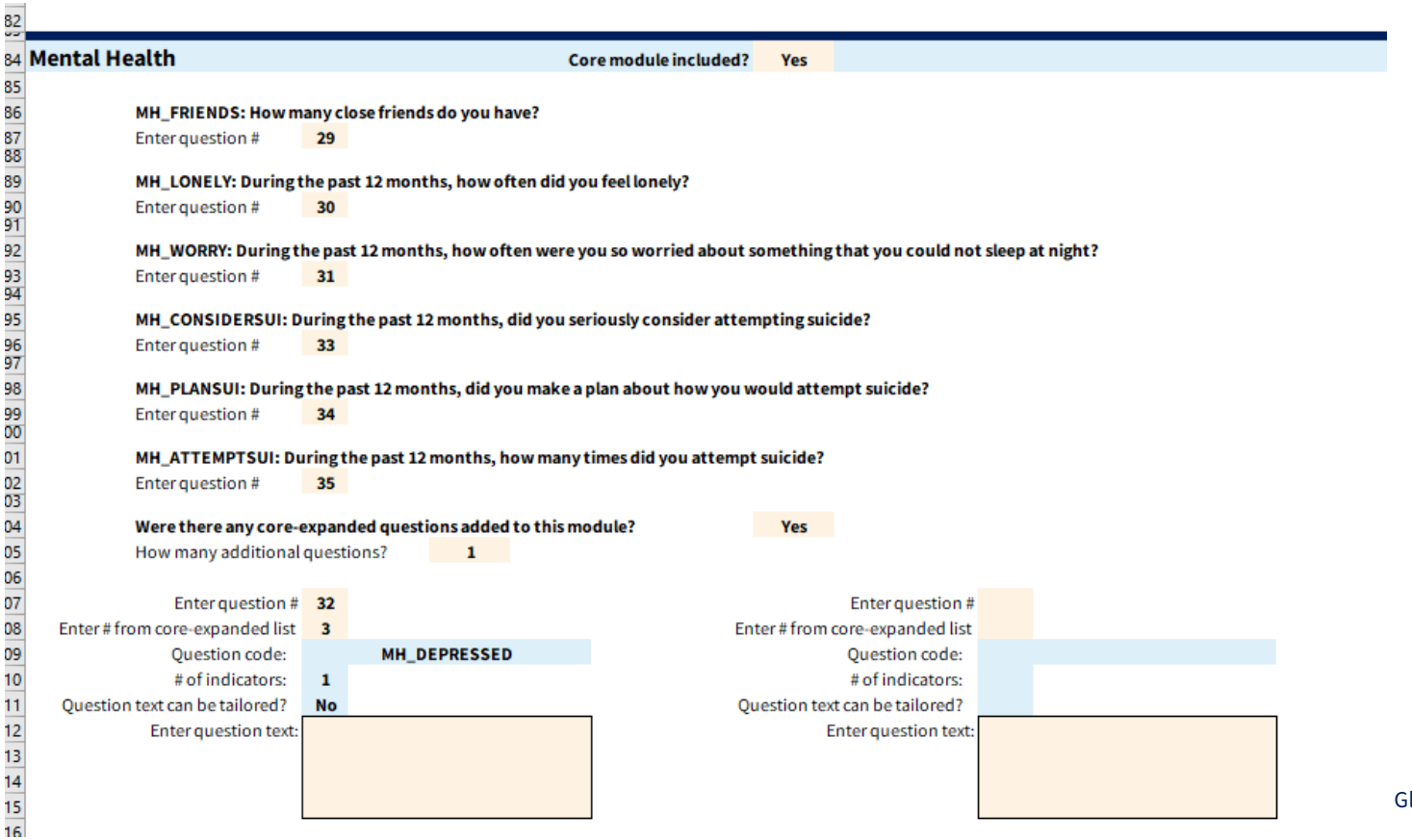
Analysis matrix (“Matrix” worksheet)

The **factsheet_section** and **factsheet_subtitle** columns are used when generating the fact sheet and indicate if the indicator is included in the fact sheet and in which section.

	A	B	C	D	E	F	G	H	I
	bin_standard	**site**	**numerator**	**denominator_resp_reduced**	**indicator_description**	**survey_question**	**var_levels**	**factsheet_section**	**factsheet_subtitle**
5	TO_B_DAYSCIG	q37	c('B','C','D','E','F','G')		TO_B_DAYSCIG: Percentage of students who currently smoked cigarettes (on at least 1 day during the 30 days before the survey)	TO_DAYSCIG: During the past 30 days, on how many days did you smoke cigarettes?	A:0 days;B:1 or 2 days;C:3 to 5 days;D:6 to 9 days;E:10 to 19 days;F:20 to 29 days;G:All 30 days	tobacco	Tobacco Use
6	TO_B_STOPCIG	q38	c('B')	c('A')	TO_B_STOPCIG: Percentage of students who tried to stop smoking cigarettes (among students who smoked cigarettes during the 12 months before the survey)	TO_STOPCIG: During the past 12 months, did you try to stop smoking cigarettes?	A:I did not smoke cigarettes during the past 12 months;B:Yes;C:No		

Analysis matrix (“Matrix” worksheet)

- There is a **matrix generation tool** available from WHO which allows you to enter information about your questionnaire and then use a macro to generate your **“Matrix”** worksheet, ensuring all formatting/labeling is correct.



82					
84	**Mental Health**		**Core module included?**	**Yes**	
85					
86		**MH_FRIENDS: How many close friends do you have?**			
87		Enter question #	**29**		
88					
89		**MH_LONELY: During the past 12 months, how often did you feel lonely?**			
90		Enter question #	**30**		
91					
92		**MH_WORRY: During the past 12 months, how often were you so worried about something that you could not sleep at night?**			
93		Enter question #	**31**		
94					
95		**MH_CONSIDERSUI: During the past 12 months, did you seriously consider attempting suicide?**			
96		Enter question #	**33**		
97					
98		**MH_PLANSUI: During the past 12 months, did you make a plan about how you would attempt suicide?**			
99		Enter question #	**34**		
00					
01		**MH_ATTEMPTSUI: During the past 12 months, how many times did you attempt suicide?**			
02		Enter question #	**35**		
03					
04		**Were there any core-expanded questions added to this module?**		**Yes**	
05		How many additional questions?	**1**		
06					
07		Enter question #	**32**	Enter question #	
08		Enter # from core-expanded list	**3**	Enter # from core-expanded list	
09		Question code:	**MH_DEPRESSED**	Question code:	
10		# of indicators:	**1**	# of indicators:	
11		Question text can be tailored?	**No**	Question text can be tailored?	
12		Enter question text:		Enter question text:	
13					
14					
15					
16					

Derived Variables ("derived_variables" worksheet)

- Derived variables are those that are generated from 2 or more other variables.
- These variables must be defined on a separate worksheet but the **indicators** to be reported for these variables should be included in the Matrix worksheet.
- The content of this worksheet is fairly technical and requires a basic knowledge of logical expressions and how to write them in R. WHO can assist you with putting together the needed code.
- On the following slides is an annotated example to demonstrate the content and structure of this worksheet.

World Health Organization

Derived Variables ("derived_variables" worksheet)

Names of the secondary variables to be derived. Our standard naming convention is to have two letters to indicate the module, then "B" for binary (data will be Yes or No) and then a short descriptive name.

Standardized names of the variables from which the new variable will be derived.
The variable names listed here must appear in the survey_question column of the Matrix worksheet.

sec_vars	req_vars	log_cond_num	log_cond_denom
DB_B_ALLSSBNONE	DB_SODA,DB_SSB	data$DB_SODA == 'A' & data$DB_SSB == 'A'	All
DB_B_ALLSSBLESS	DB_SODA,DB_SSB	(data$DB_SODA == 'A'\|data$DB_SODA == 'B'|data$DB_SODA == 'C') & (data$DB_SSB == 'A'|data$DB_SSB == 'B'\|data$DB_SSB == 'C')	All
TO_B_SMOKE_DRINK	AL_DAYS,TO_DAYSCIG	(data$AL_DAYS=='B'\|data$AL_DAYS=='C'|data$AL_DAYS=='D'\|data$AL_DAYS=='E'|data$AL_DAYS=='F'\|data$AL_DAYS=='G')	(data$TO_DAYSCIG=='B'\|data$TO_DAYSCIG=='C'|data$TO_DAYSCIG=='D'\|data$TO_DAYSCIG=='E'|data$TO_DAYSCIG=='F'\|data$TO_DAYSCIG=='G')

World Health Organization

Derived Variables (“derived_variables” worksheet)

Logical condition for the numerator using standard R coding format.

Logical condition for the denominator. If the indicator is to be reported out of all students, put “All”, else enter standard R code. In the last row below, the indicator desired is supposed to be among current smokers.

sec_vars	req_vars	log_cond_num	log_cond_denom
DB_B_ALLSSBNONE	DB_SODA,DB_SSB	data$DB_SODA == 'A' & data$DB_SSB == 'A'	All
DB_B_ALLSSBLESS	DB_SODA,DB_SSB	(data$DB_SODA == 'A'\|data$DB_SODA == 'B'|data$DB_SODA == 'C') & (data$DB_SSB == 'A'|data$DB_SSB == 'B'\|data$DB_SSB == 'C')	All
TO_B_SMOKE_DRINK	AL_DAYS,TO_DAYSCIG	(data$AL_DAYS=='B'\|data$AL_DAYS=='C'|data$AL_DAYS=='D'\|data$AL_DAYS=='E'|data$AL_DAYS=='F'\|data$AL_DAYS=='G')	(data$TO_DAYSCIG=='B'\|data$TO_DAYSCIG=='C'|data$TO_DAYSCIG=='D'\|data$TO_DAYSCIG=='E'|data$TO_DAYSCIG=='F'\|data$TO_DAYSCIG=='G')

Derived Variables (“derived_variables” worksheet)

Be sure to add a row for each derived variable in the **Matrix worksheet**. The rows for these indicators only need to contain the derived variable name in the “bin_standard” column, the value ‘Yes’ (or ‘No’) in the “numerator” column and the name of the variable plus a full text description in the “indicator_description” column. All other columns can be left blank.

bin_standard	site	numerator	denominator_resp	indicator_description	survey_ques	var_levels	factsheet_se	factsheet_sub
DB_B_ALLSSBNONE		'Yes'		DB_B_ALLSSBNONE: Percentage of students who did not drink any sugar-sweetened drink (during the 7 days before the survey)				
DB_B_ALLSSBLESS		'Yes'		DB_B_ALLSSBLESS: Percentage of students who drank any sugar-sweetened drink less than one time per day (during the 7 days before the survey)				

NOTE: *Since the indicators in the reporting documents will appear in the same order as they are listed in the matrix, it is advisable to insert any derived variable indicators before or after related indicators.*

World Health Organization