

## Example 3.A: Doer/non-doer analysis example for chlorinating drinking water

Figure 1: Data entry and division of the sample into doers and non-doers

Question number	B1	B2_1	B2_2	B2_3	B2_4	B2_sum	B3	B4	B5_3	B5_4	B6	B7
Factor	Behavior: Chlorination	Health knowledge 1	Health knowledge 2	Health knowledge 3	Health knowledge 4	Health knowledge Sum	Others' behavior	Action control	Reason 1 Good mother	Reason 2 Cheap	Respondent's Age	Monthly income
Range	(0—100%)	(0;1)	(0;1)	(0;1)	(0;1)	(0—4)	(0—4)	(0—4)	(0;1)	(0;1)	(16;80)	Natural numbers
1	91%	0	0	0	0	0	0	1	1	1	75	3200
2	88%	0	0	0	0	0	0	2	0	0	40	500
3	60%	0	0	0	0	0	1	0	0	0	21	4500
4	95%	1	1	1	1	4	2	4	1	1	27	1000
5	57%	1	1	1	1	4	1	1	0	0	18	7000
6	98%	0	0	0	0	0	3	0	1	1	37	650
7	40%	0	0	0	0	0	0	0	0	1	35	7800
8	97%	1	1	1	1	4	4	2	0	0	52	7500
9	90%	0	0	0	0	0	4	3	1	0	34	200
10	92%	1	1	1	1	4	4	3	1	1	32	600
11	55%	0	0	0	0	0	1	0	1	0	18	1600
12	90%	1	1	0	1	3	4	4	0	0	65	6000
13	78%	0	1	0	0	1	0	0	1	1	17	1500
14	64%	1	0	1	1	3	2	3	0	1	32	570
15	34%	1	1	0	1	3	3	2	0	1	34	3200
16	94%	0	1	1	1	3	3	3	1	0	44	800
17	27%	0	1	1	1	3	3	4	1	1	22	750
18	90%	1	1	0	0	2	3	4	0	1	20	10000
19	97%	1	1	1	1	4	4	3	1	0	42	200
20	32%	1	1	1	1	4	4	1	0	0	48	1500
	73%	0.50	0.70	0.45	0.70	2.35	2.35	2.00	0.50	0.50	35.65	2053.50

6) In column G, we calculated the sum score for the sub-questions B2\_1 to B2\_4 asking about different potential causes of diarrheal disease.

5) Conditional formatting function

3) Cell H8 represents the response of participant 5 to question B3, how many community members chlorinate all their drinking water. The response was 1 = less than half of them (25%).

1) Row 15 contains the responses of participant number 12.

7) Based on question B1, measuring the amount of drinking water chlorinated, we divided the sample into doers and non-doers. Doers, highlighted in green, are those chlorinating 90% and more of their drinking water. Non-doers, highlighted in orange, are those chlorinating less than 90% of their drinking water.

2) Column I represents the responses to question B4, about the extent to which the participant tries to chlorinate all drinking water.

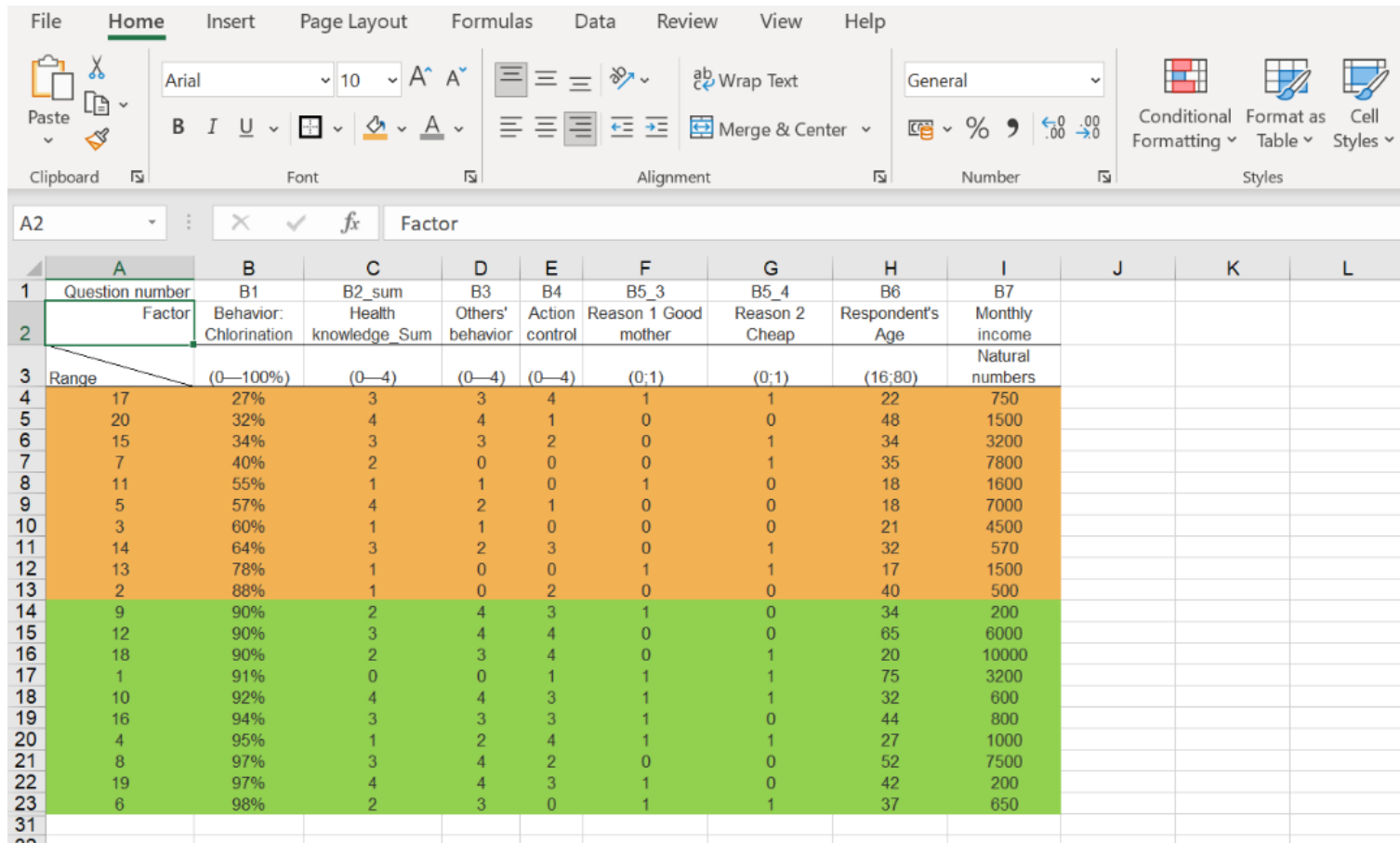
4) Column K represents the response option 2, to be a good mother, to question B5, about reasons to chlorinate drinking water. Participants who gave this reason receive a value of 1; participants who did not receive a value of 0.

*Sort the dataset according to doers and non-doers:*

Sort your data according to your main behavioral outcome variable (column B1, Behavior: Chlorination, measured in % of household's chlorinated water). In more detail, the following steps need to be performed:

- Select the whole table; selecting your data only partially will distort your data dramatically as only some columns will get sorted and others remain (!), make sure to choose the option "expand the selection" if requested.
- Make sure you have selected the option "my data has headers" within the sort command; so that your variable names will always remain in the first row
- Now sort your data according to the item that you will use to divide doers and non-doers; this could be a binomial (yes/no) or a linear measure (how much?)

For formatting cells, using colors under "cell styles" or use the "conditional formatting" options to highlight your groups.



Question number	Factor	B1: Behavior: Chlorination	B2_sum: Health knowledge Sum	B3: Others' behavior	B4: Action control	B5_3: Reason 1 Good mother	B5_4: Reason 2 Cheap	B6: Respondent's Age	B7: Monthly income Natural numbers
Range		(0—100%)	(0—4)	(0—4)	(0—4)	(0;1)	(0;1)	(16;80)	
17		27%	3	3	4	1	1	22	750
20		32%	4	4	1	0	0	48	1500
15		34%	3	3	2	0	1	34	3200
7		40%	2	0	0	0	1	35	7800
11		55%	1	1	0	1	0	18	1600
5		57%	4	2	1	0	0	18	7000
3		60%	1	1	0	0	0	21	4500
14		64%	3	2	3	0	1	32	570
13		78%	1	0	0	1	1	17	1500
2		88%	1	0	2	0	0	40	500
9		90%	2	4	3	1	0	34	200
12		90%	3	4	4	0	0	65	6000
18		90%	2	3	4	0	1	20	10000
1		91%	0	0	1	1	1	75	3200
10		92%	4	4	3	1	1	32	600
16		94%	3	3	3	1	0	44	800
4		95%	1	2	4	1	1	27	1000
8		97%	3	4	2	0	0	52	7500
19		97%	4	4	3	1	0	42	200
6		98%	2	3	0	1	1	37	650

Figure 1: Dataset after sorting according to the behavior measure, color coding for doers and non-doers

Calculate the mean scores of each behavioral factor separately for doers and non-doers

For each behavioral factor (i.e. for each question), the mean score in the responses is calculated separately for doers and non-doers. Figure 2 provides a fictional example for three psychosocial factors (*Health knowledge*, *Others' behavior*, and *Action control*), one open multiple-response question on the reasons for chlorinating drinking water, and two contextual factors (age and monthly income of the household).

Calculation and interpretation of mean scores is quite straightforward for questions with rating scales or about factors such as age; it simply means the average of responses. For yes/no questions, the mean score equals the percentage of yes responses and should be displayed in Excel as a percentage. For open multiple-response questions, we treat every response option as a separate yes/no question; 'yes' means that that response was mentioned and 'no' means that that response was not mentioned. See figure 1 and 2 for the data entry of open multiple-response questions and figure 3 and 4 for the calculation of mean scores for open multiple-response questions.

**More details:** calculating means for doers and non-doers is done with the following steps:

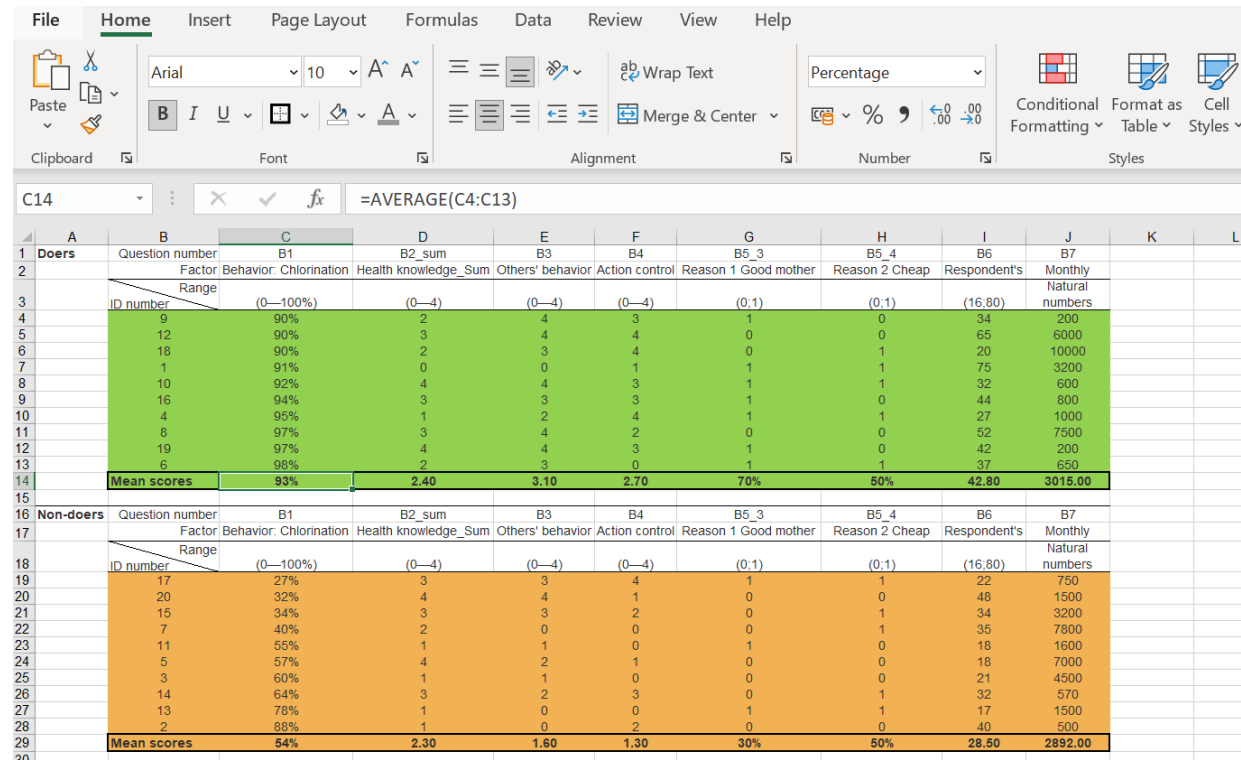
- Separate your two groups of Doers/NonDoers according to the cutoff value (yes vs. no; or a specific value that distinguishes Doers from Non-Doers for you) by inserting several extra rows between the two groups so that the data of one group lies above and the other group below that row. In this example, the cut-off point is 90%: households who chlorinate 90% or more of their drinking water are considered doers, and households who chlorinate less than 90% of their drinking water are considered non-doers.
- Calculate means for both groups separately for comparison using the arithmetic "MEAN" function
- Select one cell below the data in one variable and one of your groups
- Type "=AVERAGE(first cell:last cell)" into that cell or use the built-in function builder to do that – this should calculate the mean value of the range of cell specified
- You can now do this for all variables and both groups, or simply copy and paste the first AVERAGE cell you created into the other cells where you need the means – this will automatically update the specified cell range to that column and range
- Compare the mean scores between doers and non-doers to identify the behavior-steering factors

Next, we compare the mean scores of doers and non-doers for each behavioral factor. We calculate the differences between mean scores for doers and non-doers. The critical behavioral factors are those with the largest differences between doers and non-doers. Figure 3 and 4 provide an example of how to do so.

**More details:** Calculate differences between the means of the two groups in each variable:

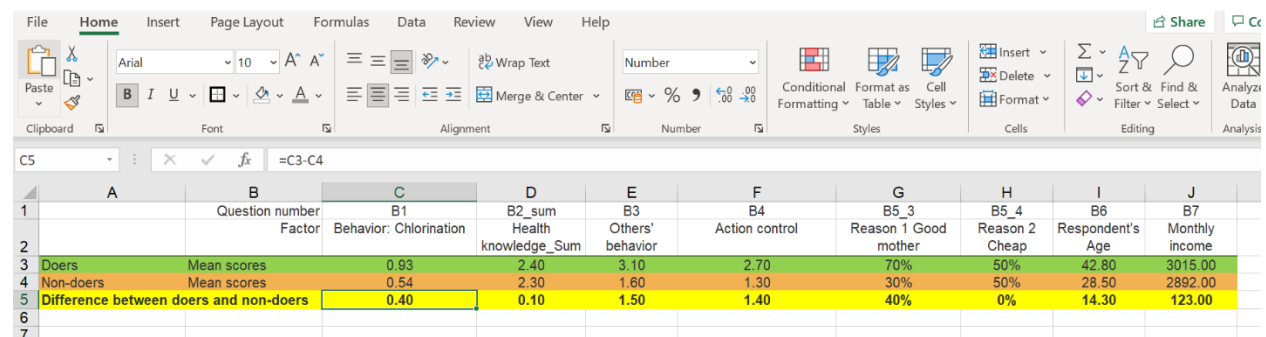
- On a new sheet, create three rows. One with Doers, one with Non-Doers and one which is named Difference between doers and non-doers. Copy-paste the mean scores from the first sheet together with the item names for doers and for non-doers. When pasting values make sure to use the option "paste values only".

- In the last row in figure 4 (marked in yellow), under the first variable, create the difference of the mean value of the doers minus the mean value of the non-doers. Repeat this, or copy-paste this cell for the rest of the variables.



	A	B	C	D	E	F	G	H	I	J	K	L
1	<b>Doers</b>	Question number	B1	B2_sum	B3	B4	B5_3	B5_4	B6	B7		
2		Factor	Behavior: Chlorination	Health knowledge_Sum	Others' behavior	Action control	Reason 1 Good mother	Reason 2 Cheap	Respondent's	Monthly		
3		Range	(0—100%)	(0—4)	(0—4)	(0—4)	(0;1)	(0;1)	(16;80)	Natural numbers		
4		ID number										
5		9	90%	2	4	3	1	0	34	200		
6		12	90%	3	4	4	0	0	65	6000		
7		18	90%	2	3	4	0	1	20	10000		
8		1	91%	0	0	1	1	1	75	3200		
9		10	92%	4	4	3	1	1	32	600		
10		16	94%	3	3	3	1	0	44	800		
11		4	95%	1	2	4	1	1	27	1000		
12		8	97%	3	4	2	0	0	52	7500		
13		19	97%	4	4	3	1	0	42	200		
14		6	98%	2	3	0	1	1	37	650		
14		<b>Mean scores</b>	<b>93%</b>	<b>2.40</b>	<b>3.10</b>	<b>2.70</b>	<b>70%</b>	<b>50%</b>	<b>42.80</b>	<b>3015.00</b>		
15												
16	<b>Non-doers</b>	Question number	B1	B2_sum	B3	B4	B5_3	B5_4	B6	B7		
17		Factor	Behavior: Chlorination	Health knowledge_Sum	Others' behavior	Action control	Reason 1 Good mother	Reason 2 Cheap	Respondent's	Monthly		
18		Range	(0—100%)	(0—4)	(0—4)	(0—4)	(0;1)	(0;1)	(16;80)	Natural numbers		
19		ID number										
20		17	27%	3	3	4	1	1	22	750		
21		20	32%	4	4	1	0	0	48	1500		
22		15	34%	3	3	2	0	1	34	3200		
23		7	40%	2	0	0	0	1	35	7800		
24		11	55%	1	1	0	1	0	18	1600		
25		5	57%	4	2	1	0	0	18	7000		
26		3	60%	1	1	0	0	0	21	4500		
27		14	64%	3	2	3	0	1	32	570		
28		13	78%	1	0	0	1	1	17	1500		
29		2	88%	1	2	0	0	0	40	500		
29		<b>Mean scores</b>	<b>54%</b>	<b>2.30</b>	<b>1.60</b>	<b>1.30</b>	<b>30%</b>	<b>50%</b>	<b>28.80</b>	<b>2892.00</b>		

Figure 2: Dataset after deviding doers and non-doers



Question number	Factor	B1	B2_sum	B3	B4	B5_3	B5_4	B6	B7	
		Behavior: Chlorination	Health knowledge_Sum	Others' behavior	Action control	Reason 1 Good mother	Reason 2 Cheap	Respondent's Age	Monthly income	
3	Doers	Mean scores	0.93	2.40	3.10	2.70	70%	50%	42.80	3015.00
4	Non-doers	Mean scores	0.54	2.30	1.60	1.30	30%	50%	28.50	2892.00
5	Difference between doers and non-doers		0.40	0.10	1.50	1.40	40%	0%	14.30	123.00

Figure 3: Mean scores for doers and non-doers

For open multiple-response questions, we have to compare each response option between doers and non-doers. When a question has many response options, this involves a great deal of effort, and one can quickly lose track of the comparisons. Therefore, we recommend measuring as many factors as possible by closed questions with rating scales (see Step 2.1).

In Figure 4, the difference in psychosocial factors between doers and non-doers is smallest in *Health knowledge* (0.10), larger in *Action control* (1.40), and largest in *Perceived others' behavior* (1.50). This means that *Others' behavior* is most critical, followed by *Action control*. When we examine the reasons mentioned for chlorinating drinking water, there is a large difference (40%) in reason 3, to be a good mother, which is much more frequently mentioned by doers than by non-doers, and no difference in reason 4, because chlorination is cheap (0% difference). Therefore, *Others' behavior* and *Action control* should be targeted through BCTs as well as being a good mother. In the contextual factors, doers and non-doers differ in age (doers are on average 14.30 years older than non-doers) but only marginally in their households' monthly income (123 Kenyan Shilling). Of course, we cannot change participants' ages. However, we can tailor our interventions to the critical age group, in this case young adults. The results of this example are summarized together with the potential interpretation of the results in table 1: Interpretation of results.

Table 1: Interpretation of results

Factor	Item / Question	mean value	difference	interpretation	decision
Behavior: chlorination	How much of your household's drinking water do you chlorinate?	Doer: 93% Non-Doer: 54%	40%	Doers chlorinate 93% of their drinking water on average, whereas non-doers only chlorinate 54% of their drinking water.	selected for the intervention
health knowledge	I will present you some potential causes of diarrhea. Could you please tell me for each whether it is a cause of diarrhea or not?	Doer: 2.4 Non-Doer: 2.3	0,1	Doers and non-doers do not differ strongly in health knowledge.	not selected for the intervention
other's behavior	How many people of your community chlorinate all their drinking water?	Doer: 3.1 Non-Doer: 1.6	1,5	Doers perceive more that others in their surrounding also chlorinate their drinking water.	selected for the intervention
action control	How keenly do you try to chlorinate all your drinking water?	Doer: 2.7 Non-Doer: 1.3	1,4	Doers more keenly try to chlorinate all their drinking water.	selected for the intervention
reason 1: good mother	What are your reasons to chlorinate your drinking water?	Doer: 70% Non-Doer: 30%	40%	70% of the doers mention that a reason why they chlorinate their drinking water is because they want to be a good mother, only 30% of the non-doers say so.	selected for the intervention
reason 2: cheap	What are your reasons to chlorinate your drinking water?	Doer: 50% Non-Doer: 50%	0%	Doers and non-doers do not differ on the frequency of mentioning the reason for chlorination that it is cheap.	not selected for the intervention
respondents age	How old are you?	Doer: 42.8 Non-Doer: 28.5	14,3	Doers are on average older than non-doers.	considered in the intervention

monthly income	What is the monthly income of your household?	Doer: 3015 Non-Doer: 2892	123	Doers and non-doers do not strongly differ on their monthly income.	not considered in the intervention
----------------	---	------------------------------	-----	---	------------------------------------

Note that a doer/non-doer analysis was essential to determine the critical behavioral factors; a simple calculation of the mean scores in the population would have yielded other, potentially misleading, results. In this instance, examining the mean scores in the population could have led to the conclusion that *Health knowledge* was the most critical to target, as *Health knowledge* is quite low (see cell G32, bordered in violet in figure 1). However, the doer/non-doer analysis shows that doers and non-doers differ only marginally in *Health knowledge* (see figure 4 and table 1). In other words, *Health knowledge* cannot explain why some people chlorinate their drinking water (doers) while others do not (non-doers). Thus, *Health knowledge* is not a critical behavioral factor and should not be prioritized in an intervention.