**Guide to the WASH in Healthcare Facility Survey and Evaluation R Tool**

If you have questions, email stephenpathilton@gmail.com

Table of Contents

[Summary 2](#_Toc75782878)

[Section 1: What is in the WASH in HCF RTool? 3](#_Toc75782879)

[Section 2: How to Use the Tool 4](#_Toc75782880)

[Data cleaning 4](#_Toc75782881)

[Downloading and setting up R and RStudio 5](#_Toc75782882)

[What Lines Must be Changed? 5](#_Toc75782883)

[Running the code 6](#_Toc75782884)

[Results 7](#_Toc75782885)

[Section 3: How the Tool Works 9](#_Toc75782886)

[Section 4: Common Questions and Editing the Script 13](#_Toc75782887)

[Common Issues 13](#_Toc75782888)

[Minor Changes to the Script 14](#_Toc75782889)

[Major Changes to the Script 16](#_Toc75782890)

# Summary

This R Tool was created as part of the Dicastery for Promoting Integral Human Development’s project to support 150 Catholic healthcare facilities in evaluating and address water, sanitation, and hygiene deficits. It has since been adapted so that it could be used in other healthcare facilities (whether secular or faith-based). It is meant to quickly assess whether healthcare facilities meet basic water, sanitation, and hygiene standards. It also generates reports for each healthcare facility. These reports highlight strengths, issues that may be addressed without much additional funding (minor issues to resolve), and issues that will likely require additional funding (major issues to resolve).

This is a companion tool to the *HCF Survey and Evaluation* Excel Tool. That tool is easier to use, but only allows assessment of up to 20 facilities at a time. This tool allows assessment of more healthcare facilities, and automatically generates facility reports.

Previous experience with R or coding is not required to use this tool. This guide will highlight all necessary steps the user must take and will provide a general overview of the steps this code takes.

This guide is organized by how much the user wants to know about the tool.

* The first section highlights what files are in this tool.
* The second section provides a step-by-step guide on using the tool. **This and the first sections are the only that you need to read to use this tool.**
* The third section provides information on how the tool works.
* The fourth section provides answers to potential issues and provides instructions for adding to the assessment.

This guide was written using a windows computer. The process for Mac is likely very similar if not exactly the same.

# Section 1: What is in the WASH in HCF RTool?

The First folder has:

* HCR Survey and Evaluation RTool.Rmd
  + This is the file/script that is used to analyze data and generate reports. This file is operated using the program RStudio.
* WASH in HCF Evaluation Template.xlsx
  + I recommend that you fill this document with your survey information. It is preformatted to work with the R script.
* WASH in HCF Evaluation Template filled anonymous catholic.xlsx
  + This is the template filled with the information of some anonymized healthcare facilities. It can be used for a trial run of the script, which the script is already setup to use.

The *Classification and Prompt Data* subfolder has:

* Classification Answers.xlsx
  + This Excel document list which survey answer fall into the three categories for this assessment (strengths, minor issues to resolve, and major issues to resolve).
* Prompt Answers.xlsx
  + This Excel document lists which prompts will be output on facility reports if certain conditions are met.

The *Facility Report Output* is the suggested folder for the R script to output facility reports and summary statistics. It also has:

* WASH Survey Gap Recommendations.pdf
  + This document suggests next steps for healthcare facilities depending on the results of their survey. This should be sent to facilities along with their facility reports.

# Section 2: How to Use the Tool

To use this tool, one must ensure their data are properly formatted, setup R and RStudio (if not already done), change file paths, and then run the code. If you want a trial run before using collected survey data, you can use *WASH in HCF Evaluation Template filled anonymous catholic.xlsx*.

## Data cleaning

The survey data must be properly formatted for the tool to recognize the survey answers and to properly classify them. Not doing so can result in incomplete or blank reports, and improperly classifying which healthcare facilities meet basic standards.

If you conduct the survey on *WASH in HCF Evaluation Template.xlsx,* then it should be formatted correctly. You can also paste data from the *HCF Survey and Evaluation* Excel Tool, but you will need to add questions i, ii, and iii.

Common formatting issues include:

* Typing the answers: If the answers are not typed *exactly* as they are found in *Classification Answers.xlsx,* the code will not be able to recognize and classify survey answers. For example, if any survey answers from numbered questions have commas, the code will not recognize it.
  + To prevent this issue, I recommend using *WASH in HCF Evaluation Template.xlsx* or *HCF Survey and Evaluation Tool.xlsx* which have drop down menus.
* Not converting question 6 to days per year: This must be input just as a number or else it will not be recognized. For example, if water runs continuously one would input “0”.
* Having healthcare facility names with characters that cannot be saved: Some special characters used can prevent the filename for reports from being valid. This [link](https://www.mtu.edu/umc/services/websites/writing/characters-avoid/) suggests characters to avoid using in the “Health Facility Name, for File:” line.
* Having too long of a path name: Windows only supports path names (noting where a file is saved) of 255 characters. Outputting reports into nested subfolders can take up too many characters. Using a short name into the row title “Health Facility Name, for File:” *WASH in HCF Evaluation Template.xlsx* in can help.
* Inputting on columns without healthcare facilities: DO NOT type anything on columns without healthcare facilities. If left blank the script will recognize that those are not healthcare facilities. If some cells are filled they will be loaded, which could lead to an error or lead the script to think that there are a different number of healthcare facilities.

Two easy ways to check whether the data are properly formatted are:

1. Use the *WASH in HCF Evaluation Template.xlsx* file and follow its prompts.
2. You can paste answers from some healthcare facilities into the Excel Evaluation tool (*HCF Survey and Evaluation Tool.xlsx*). If correctly formatted, many cells should be colored green, yellow, and/or red on the “Input Color Coded” sheet. Questions M, 1, 3, 4, 8, 9-23, and 26 nearly always should be colored (unless other is selected), and other will be as well.

Calendar

Description automatically generated with low confidence

*Figure 1: Demonstration of using the Excel sheet to check if answers have been properly typed in. The example on the left has and has many more colored cells. In the example on the right, similar (but not exact) answers were typed out, and the tool did not recognize them.*

## Downloading and setting up R and RStudio

This tool uses the R coding language, and you will access it on RStudio IDE (desktop version). Both are free. R can be accessed here, and [RStudio](https://www.rstudio.com/products/rstudio/download/) can be accessed here. This [link](https://rstudio-education.github.io/hopr/starting.html) provides instruction on installing both.

## What Lines Must be Changed?

Before running the program, you must make certain changes to the code. These allow the code to recognize which Excel documents to use for survey data, classifying answers, and prompts to output, as well as to know where to save reports.

The only lines that need to be changed are all in the one “chunk” (section of code) shown below. By saving variable names with path names here, later on the script will use them to input and output information.

Text

Description automatically generated

*Figure 2: This shows what code needs to be changed. The section highlighted in yellow is the HCF survey data. The section highlighted in orange is the classification answers (determines what category a survey answer belongs in). The section highlighted in blue is the prompts that will be output on the report. The section highlighted in green is the excel document that will be output that shows how many facilities meet basic standards. The section highlighted in gray is the excel document that will be output that shows which facilities meet basic standards. The section highlighted in red is the folder where you wants the reports to output.*

For lines 42, 43, and 44, you need to replace the highlighted code with the paths (including file names) of their survey data, their classification answers, and the prompt answers.

For lines 45 and 46, the users need to replace the highlighted section with where they want the basic standards excel documents to output, as well as the file names. You can keep the file names *Numbers\_Facilities\_Basic* and *Which\_Facilities\_Basic* after the final forward slash if desired.

For line 47, replace the highlighted section with the path for the folder that you would like to put the facility reports in. The code will automatically create file names.

*Hint: One easy way to find the file path is to copy and paste the title bar on File Explorer (on the left of the search bar). From there, change all backward slashes to forward slashes, and then add your file name (for survey\_data\_fn, answer\_class\_fn, and prompt\_answers\_fn).*

Graphical user interface, text, application, email

Description automatically generated

*Figure 3. The red square shows one easy way to find a file path name. From there, you would need to replace all backslashes with forward slashes. If this is for a line that grabs a file, a forward slash and the filename would be added at the end.*

## Running the code

To run the code, click the green arrows on lines 41, 54, 69, 418, and 436. Each will take some time (~30 seconds) so wait between each click.

Text

Description automatically generated

*Figure 4: Clicking the green arrow will run all code in the grey box (lines 54-63), which in this case downloads R packages that allow this script to work.*

Table

Description automatically generated

*Figure 5: An unfilled green bar indicates that the section of code is still running. Do not run another section until this bar goes away.*

There are two common errors to avoid when running the program.

1. If you need to run the script again, you must shut down and reopen RStudio. One R Package used for the report (at the end) overwrites one used for the analysis (at the beginning).

*Figure 6: A package used for writing reports overwrote one needed to load the data.*

1. Make sure that reports do not go to a folder with the same file names. If this occurs, RStudio may shut down.

## Results

If the script runs successfully, you should have the following:

* *Number\_Facilities\_Basic.xlsx*: lists how many facilities met basic water, sanitation, and hygiene standards, and how many healthcare facilities were assessed.
* *Which\_Facilities\_Basic.xlsx*: shows which healthcare facilities meet these basic standards.
* The facility reports, which should list general information about the healthcare facility, its strengths, its minor issues to address, and its major issues to address.

Text

Description automatically generated

*Figure 7: An example of an anonymized report. In your reports, “HCF #10” would be replaced with the name of the healthcare facility.*

I recommend emailing each facility their report and the decision tree (*WASH Survey Gap Recommendations.pdf).*

# Section 3: How the Tool Works

The script was designed with a goal of user friendliness (though those with experience coding may notice it was not designed to run as efficiently as possible). The diagram below provides a high-level summary on what the script does. Additional information is provided below.

Diagram

Description automatically generated

*Figure 8: Diagram representing what the script does and which lines complete each task. Yellow boxes represent package installation, blue boxes represent computation, and green boxes represent outputs. The first box is the only one the user needs to edit on the RScript.*

### User Inputs File Names

This is the only code that the user needs to adjust. They will add in path and file names for the survey data, the Excel documents with answer classification information and prompts, and where they want reports and summary data to be saved. The variables created here are used later in the code.

### Installing Necessary R Packages

To complete the analysis, the packages *readxl, writexl, tidyverse, forcats,* and *dplyr* are needed. Additional packages will be needed later to produce the report documents. These document-producing packages partly overwrite the packages required for data analysis, which is why RStudio must be restarted to rerun the script.

### Formatting Data

The script removes certain empty rows and columns to make later processing easier. This makes it easier to operate the loops that classify what category a survey answer belongs in and to assign prompts.

### Classifying Answers

For each survey question, the script checks each healthcare facility’s answers. It references *Answer Classification.xlsx* to determine if each survey answer falls into the strengths, minor issues to resolve, and major issues to resolve categories. If it does, a number is assigned to indicate which category it falls into.

Chart, box and whisker chart

Description automatically generated

*Figure 9: This categorization is done using ‘for loops’ and ‘if loops.’ It starts at the first question on the first healthcare facility and checks what category the answer belongs in. If the survey answer falls into a certain category, it assigns a numerical code. It then checks the first question on the next healthcare facility (for each healthcare facility). After that question has been checked in all healthcare facilities, it then goes to the next question (for each question) and repeats the process.*

### Evaluating if Basic WASH Standards are Met

Using the classified answers, the script determines if each healthcare facility meets the standards for basic water, sanitation, hygiene, waste management, and environmental cleaning. The definitions for each can be found in this [link](https://washdata.org/monitoring/health-care-facilities). This also uses two ‘for loops’ and an ‘if loop.’ It also formats these answers for Excel.

### Outputting the Basic Standards Analysis Excel Sheets

It outputs two Excel sheets. *Number\_Facilities\_Basic.xlsx* lists how many facilities met basic water, sanitation, and hygiene standards, and how many healthcare facilities were assessed. *Which\_Facilities\_Basic.xlsx* shows which healthcare facilities meet these basic standards. The user can change the names of these documents in the *User Inputs File Names* section if desired.

### Removing Redundant Answers

There are multiple survey questions around water, sanitation, and running water in maternal wards. In order to avoid too much repetition on the reports (e.g. *Having a Safe Water Source* and *Having a Water Source on Premises*), in some cases single prompts covering multiple answers were provided. The script first flags which healthcare facilities have this repetition, and then later deletes and replaces prompts.

### Assigning Prompt Based on Classification

Depending on how each survey answer is classified, different prompts are assigned. This is done using the numbering system on the Classifying Answers sections and the prompts on the *Prompt Answers.xlsx* Excel sheet.

### Consolidating Prompts

There are many gaps between prompts. This removes these from each healthcare facility so prompts can be output without blank lines on the report.

### Installing Document Creation Packages

The *officer*, *flextable*, and *magrittr* R packages are needed to output reports. Formatting styles are also selected for the reports. These packages partly overwrite the packages required for data analysis, which is why R must be restarted to rerun the script.

### Document Creation Loops

The remaining subsections are all part of the document creation loop. One iteration is needed per healthcare facility. In each iteration, data from that healthcare facility are selected, the report is formatted, the prompts are output, and the document is saved. Then another loop begins with the next healthcare facility.

*Hint: Be sure that the folder the reports are being saved into are blank. If there are reports with the same name, RStudio may shut down.*

### Select Data from a HCF

Variables for the individual healthcare facility are made with survey answers, prompts, and other important information. This is later used to provide identifying information on the report.

### Format Report Document

These lines set the title, subtitles, and report excerpts. Identifying information (e.g. the facility name and address) is added to the report. Formats for text are assigned. If you want to change some of the text (e.g. title or paragraph describing the Dicastery Initiative), this is where you would do so.

### Outputting Strengths, Minor Issues to Resolve, and Major Issues to Resolve Prompts

Subloops are used to go through the consolidated prompts and output them. The strengths, minor issues to resolve, and major issues to resolve are each done separately. These lines also output the subheader and caption for each section. If a facility does not have any prompts of a certain category, no bullets, subheaders, or captions are output for that section. For example, if a facility has no *Minor Issues to Resolve* Prompts, then no bullet points or subheaders for *Minor Issues to Resolve* will appear in the report document.

### Outputting Report Document

These lines save each report document. Each is saved as *folder path\_healthcare facility name\_WASHinHCF Report.docx*

*Hint: Be sure that the folder the reports are being saved into are blank. If there are reports with the same name, RStudio may shut down.*

*Hint: If you want to have an organization name as part of the name file, refer to section 4.*

# Section 4: Common Questions and Editing the Script

This section provides guidance for common issues that can prevent the script from running. In addition, it provides guidance on how to make certain changes to the script. Some of these are relatively simple, such as changing preset text on the reports (e.g. the paragraph describing the purpose of the report). Others may be more challenging, such as adding questions to the survey.

## Common Issues

### What if the output seems incorrect?

If the report looks incorrect (e.g. it has few prompts, they do not seem to match), there are a couple potential causes.

A common issue is that some Excel sheet, likely the survey data, is improperly formatted. This could either be an error with dimensions (which may prevent the tool from comparing the survey responses to the correct classification answers), or with the survey answers themselves (preventing the tool from recognizing the answers).

One way to check for errors for the former issue is to check the number of rows (observations) and columns (variables) in the Global Environment. *hcf\_data* should have 47 rows/obs. and as many columns/variables as there are healthcare facilities. *Answers\_Major, Answers\_Minor,* *Answers\_Strengths, responses\_major, responses\_minor, and responses\_strengths* should each have 35 rows/obs. One common cause that changes the dimensions is accidentally typing in a cell that should not be typed in.

Graphical user interface, text, application, email

Description automatically generated

*Figure 10: The “Global Environment” is a place where you can check the dimensions of the variables. The red oval shows the number of row, and the blue oval shows the number of columns (in this case healthcare facilities).*

An easy way to ensure that the survey answers are recognizable is to paste answers from some healthcare facilities into the *HCF Survey and Evaluation Tool.xlsx*, and ensure that cells in the *Input Color Coded* sheet are colored (refer to Figure 1).

### Why can’t I get it to rerun?

If you run the full script, and for some reason need to rerun it (e.g. you realized that you needed to reformat your survey sheet), you must restart R. One of the packages needed to create the reports overwrites code needed to load the survey data.

### Why am I getting the “R Session Aborted” error?

One possible cause is that you already have file and pathnames that are the same as the generated reports when you run the script. Delete those files to retry.

### Why am I getting the “Error in dirname(epath) : path too long” error?

A pathname can only be 255 characters. With long healthcare facility names, or if the reports are being saved in a path with many subfolders and/or long subfolder names, this limit may be exceeded. For this reason, the organization name was removed as part of the filename.

If you encounter this error, try saving in a folder with a small pathname or shortening the HCF file name in the Excel Sheet.

### What if my computer does not have enough memory to run this script?

This script should not be too computing intensive, so it is unlikely that you will have an error running it for this reason. If you do, try shutting down other programs (especially your internet browser) or only running fractions of your healthcare facility surveys at a time (e.g. 50 facilities instead of 200).

### Other

On some occasions, the script did not run when certain cells were left blank. The problem may have been fixed. If the script will not run, I would recommend typing an answer in the blank cells that does not correspond with an answer. For example, in the delivery room questions, on the filled template I input “Did not answer” on what used to be blank cells.

## Minor Changes to the Script

### What if I want to change some of the text in the report?

The title of the report, the italicized paragraph describing the purpose, and the descriptions below *Strengths, Minor Issues to Resolve,* and *Major Issues to Resolve* were selected as part of the Dicastery for Promoting Integral Human Development’s initiative to assess 150 healthcare facilities. You can change all of these.

To change the title of the report, you will need to adjust the highlighted section of line 449 shown below.

A picture containing text, screenshot, indoor

Description automatically generated

*Figure 11: To change the title of the report, replace the text within the quotation marks with what you would prefer it to say.*

To change the introductory text of the report, you will need to adjust the highlighted section of line 487 shown below.

Text, letter

Description automatically generated

*Figure 12: To change the description in the introductory paragraph, replace the text within the quotation marks with what you would prefer it to say.*

To change the descriptions below *Strengths, Minor Issues to Resolve,* and *Major Issues,* you can adjust liens 507, 524, or 541. The Figure below shows what to change on the strengths section.

Text

Description automatically generated

*Figure 13: To change the description below the subheaders, replace the text within the quotation marks with what you would prefer it to say.*

If you want less identifying information on each healthcare facility’s report (e.g. the Director’s Name), you can put a hash sign in front of the lines of code that you do not want.

Text

Description automatically generated

*Figure 14: To remove any of these topics from the reports, put a hash sign in front of “body\_add\_fpar.” This cancels out that line of code. For example, if you did not want to have births per month on the report, you would put a hash sign on line 480.*

### What if I want to organize healthcare facility names by organization?

A group analyzing healthcare facility data may have multiple organizations reporting to them. To help organize the reports, it may be helpful to have the organization name at the beginning of the file (*org name\_HCF name\_WASHinHCF.docx).* The script was originally written using this convention, which was changed later so that path names would not require too many characters.

To make this change, replaces line 554 with:

file\_name = paste(reports\_fn,responsible\_org,”\_”,hcf\_name\_file,"WASHinHCF",".docx")

## Major Changes to the Script

### How do I add a question?

You may decide that you would like to add questions to the survey. If you do so, you will have to make multiple changes to the script and associated files.

Step 1) Adding the question on the survey

To start, you will need to add the questions onto the excel survey document (*WASH in HCF Evaluation Template.xlsx).* While you can add this new row or rows anywhere, I strongly recommend doing so after the delivery room questions as this will reduce how much you need to change the R script. It will also ensure that questions remain the same number, which can be helpful if this survey is conducted multiple times. Even if you decide that the new questions belong elsewhere during the survey, moving them to the end for the sake of running the script will make the process much easier.

Table

Description automatically generated

*Figure 15: While new questions can be added anywhere, adding them at the end (line 55) requires the least changes.*

I recommend using a dropdown menu for the possible survey answers. This will require less data cleaning afterwards. To do so, click on a cell where you would like a user to input an answer (e.g. D55). Click Data Validation (shown below).

Graphical user interface, application, table

Description automatically generated

*Figure 16: What to click to find Data Validation.*

Select list. Type each option separated by a comma. These will then appear as selectable options.

*Graphical user interface, application

Description automatically generated*

*Figure 17: How to setup a drop-down list in an Excel cell.*

After you have the first cell setup, you can then paste it across the row (for the other healthcare facilities).

Step 2) Change the associated excel documents

For the script to work, the document that declares which survey answers fall into the *strengths, minor issues to resolve,* and *major issues to resolve* categories (*Classification Answers.xlsx* ) and the document that assigns prompts for the reports (*Prompt Answers.xlsx*) must match. Each document has a tab for each category.

For both, you must make sure that the questions follow the same sequence. That is, if you added the survey question to the end, add it to the end for all tabs. If it is added between two questions, it must be between the same two questions. Make sure that the acceptable answer matches *exactly* with the drop down.

I recommend that you save all documents (*WASH in HCF Evaluation Template.xlsx, Classification Answers.xlsx, Prompt Answers.xlsx,* and the RScript) with a new name.

Step 3) Change the script

The first changes on the script that must be made are to ensure that the file paths grab the updated files. Line 42 must be changed for *WASH in HCF Evaluation Template.xlsx,* line 43 must be changed for *Classification Answers.xlsx,* and line 44 must be changed for *Prompt Answers.xlsx.*

If you added new survey questions after all questions, you will not need to adjust the code for data cleaning, determining which healthcare facilities meet basic standards, and removing redundant answers. In which case, the script should be able to run your new version of the survey.

If you integrated the survey questions before the end, you will need to adjust the row numbers in lines 80, 81, 169-205, and 272-306. Once you adjust lines 80 and 81, using the variable hcf\_data\_test will be helpful for identifying which rows correspond with questions of interest.