

Calibration Method for Campaign Costing Studies

R script guide, September 2021

Purpose: The goal of this script is to provide a walkthrough of how the sampling weights were adjusted via the calibration method using R.

#Code developed by ThinkWell with support from Emma Clarke-Deelder and Ryoko Sato (Harvard T.H. Chan School of Public Health). This was developed as part of the Immunization Costing Action Network (ICAN) project, which was supported by the Bill & Melinda Gates Foundation.

#Please visit our website to learn more:

<https://immunizationeconomics.org/ican-campaign-costing>

<https://thinkwell.global/projects/building-country-capacity-immunization-delivery-costing-evidence-use/>

#Calibration is used to test whether the sampled sites were an accurate representation of all sites that participated in the campaign. The calibration method adjusts the inverse probability of sampling weights using the volume of vaccine doses delivered at each site and the total delivery volume for the entire campaign. The calibration method was performed using the raking function as seen below.

First, load the necessary R packages to be used

```
library(plyr)
```

```
library(dplyr)
```

```
library(tidyr)
```

```
library(tibble)
```

```
library(openxlsx)
```

```
library(survey)
```

###-----Import the dataset to be calibrated-----

#Note that the calibration dataset must contain a) the names (or site numbers) of the data collection sites; b) the inverse probability of sampling; and c) the delivery volume of each data collection site. This guide will use the sample dataset that was used to walk through the the weighted average unit cost calculation

```
data_sample <- read.csv(file = './Calibration_sample.csv') #Note that you have to specify the file path where the calibration dataset is located
```

#Now create a vector (i.e. a sequence of numbers in R to be used later) that contains a) the total number of data collection sites or health facilities within the study population (1,324 in row 25); and b) the total delivery volume (i.e. total vaccine doses delivered) for the entire campaign (4,245,540 in row 25). These will be used later on in the calibration function

```
totalVolume = c(1324, 4245540)
```

###-----Establish the survey design for the svydesign function in R-----

#The svydesign package in R is commonly used to analyze sampled survey data. We will be using the calibrate function within this package for this calculation. In this case, set "id" to the facility number variable, "weights" to the inverse probability of sampling variable, and "data" to the name of the data frame containing the calibration dataset

```
design <- svydesign(id = ~Facility_number, weights = ~inverseWeight, data = data_sample)
```

#Now use the calibrate function to adjust the sampling weights using the total delivery volume. For this, you will need to calibrate against a known total delivery volume of vaccines for the entire campaign while using the "raking" function

```
design_Cal_raking <- calibrate(design, ~ totalDoses, totalVolume, calfun = "raking")
```

#Now extract the calibrated weights from the raking function as per below

```
calibrated <- as.data.frame(design_Cal_raking$postStrata[[1]]$w)
```

#Now create a data frame that contains the list of facility numbers to start creating an updated table with the calibrated sampling weights

```
facilityNumbers <- as.data.frame(data_sample$Facility_number)
```

#Now combine the facility names and the calibrated weights data frames together to create a clean table while renaming the variable names accordingly

```
calibratedWeights <- cbind(facilityNumbers, calibrated) %>%
```

```
rename(facilityName = `data_sample$Facility_number`, # rename the variable names
```

```
calibratedWeight_Raking = `design_Cal_raking$postStrata[[1]]$w`)
```

#Now view the combined table with the calibrated weights. You can now use these in place of the original sampling weights when calculating the unit cost

```
view(calibratedWeights)
```