**ASSESSMENT: IMMUNIZATION P-3 POST, DAPM**

**Instructions:**

* **You have** 3 hours **to answer all questions**. Tests received after the time limit will not be accepted. Please use your time effectively and try to answer all questions.
* **This test is confidential.** Please do not share any further.
* **Plagiarism or seeking outside assistance will result in automatic disqualification.** Please respond to all questions in this test using your skills and experience.
* **Tests will be graded anonymously.** Please do not write your name or any identifiable information on the test.
* **All data manipulations should be completed using code.** Do not make any edits directly in the Excel sheet.
* **All codes and responses to the written answers should be submitted via email to Sandy Chien Yi Chuang at** [**scychuang@unicef.org**](mailto:scychuang@unicef.org)**.**
* **In addition to email submission,** candidates also have the option to submit their work through GitHub. Kindly share the appropriate GitHub link with **bquere**.

**Part 1 (use Excel DHIS2\_data\_Ghana\_v1)**

You received the following Excel from Ghana (DHIS2\_data\_Ghana\_v1) consisting of three sheets. The first sheet (Service\_data\_1) has data for maternal health services, the second sheet (Service\_data\_2) has data on immunization and the third sheet (Service\_data\_3) has mortality data.

1. Combined the three service data sheets to form one analytical data set. Summarize any challenges identified during the processing and combining of the three sheets and how the challenges were addressed to allow for successful merging. The code used to process and combine the three sheets must be submitted along with your answer.

Before merging/joining data, it’s best to do some data quality checks on the columns used to join. Firstly I assessed missingness. One of the districts in the mortality data was missing a name – I identified which district was missing by doing a comparison with the other two data sheets (maternal and immunization) using an anti-join to see which was missing in Mortality. This was Guan. I then replaced the missing District name with ‘Guan’ and added the months and years, so that there were 60 rows for Guan.

There were 10 completely missing rows in the mortality data (ie. They contained no data). Once I had fixed Guan, I filtered out these missing observations using filter(!is.na(District)). Now the number of observations match across the three datasets.

There are often issue with region names being spelled differently or inconsistently containing accents. I checked for accents in any names across the three data sheets. There were 4 observations in the mortality data where Keta had an accented e. I fixed the accented ‘e’ to a non-accented ‘e’ for consistency with all the other data.

It is always good to check for leading and trailing white space in data label names. There was leading white space in District names. I removed leading and trailing blank spaces to ensure correct joins by creating a format to remove white space and passing District column through the format.

I was able to do the join after this and validated by comparing number of non-missing values with the input data.

**Part 2 (use Excel ghp3\_2)**

You are provided with the following data from Ghana to analyze and provide recommendations to the Ministry of Health.

Below is the data dictionary

year: year

country: name of the country

admin1: regions in the country

admin2: districts in the country

doses: number of children who received the given antigen

target\_number: number of children **expected** to receive the antigen

target\_define: description of the target number of children expected to receive the antigen

vaccine name: name of the antigen

Based on the data provided for part 2:

2a. How many regions do we have in Ghana

Regions are admin1 regions. In 2018, there are 10 regions. In 2017-2022, there are 16 regions.

2b. How many districts do we have in Ghana

Districts are denoted by admin2 levels. The number of districts by year are as follows:

* 2018 = 216 districts
* 2019-2021 = 260 districts
* 2022 = 261 districts

2c. Which region has the largest number of districts in the country

In all years, Ashanti has the largest number of districts in the country.

In 2018, Ashanti had 30 districts and in 2019-2022, Ashanti had 43 districts.

In Ghana, BCG is recommended to be given at birth. Diphtheria tetanus pertussis (DTP) is a three-dose series vaccine recommended to be given at ages 6 weeks, 10 weeks, and 14 weeks. Unlike other vaccines, a unique thing about the DTP series is that a child cannot receive DTP3 without receiving DTP2 or receive DTP2 without receiving DTP1. The first dose of Measles-containing vaccines (MCV1) and Yellow Fever Vaccines (YFV) is recommended to be given at 9 months.

2d. Briefly summarize the quality of the data received with respect to **consistency between the related service indicators** at the **admin1 level** from **2018 to 2022**.

First I calculated weighted vaccination coverage at the admin1 level for all admin1-vaccine-years 2018-2022. I notice that coverage is greater than 100% for many admin1 regions. I also checked that that the target population across each region and related antigen are the same over the years.

The data has number of doses. Of DTP1 and DTP3. Number of DTP1 doses should be greater than number of DTP3 doses as children are more likely to receive the first dose than they are to complete the series. When a child receives DTP1 and not DTP3, this is referred to “ DTP drop-out”. Across the dataset, there are 41 observations where number of DTP3 doses is greater than DTP1 doses. This is the case for some admin1 regions for all years 2018-2022, such as Greater Accra. When calculating weighted admin1 region coverage (which is better to look at than absolute numbers as they contain noise) 39 regions have DTP3 coverage greater than DTP1 coverage when rounding coverage to a whole number.

As MCV1 and YFV are recommended to be given at the same time, we would expect MCV1 and YFV coverage track relatively closely and have similar coverage. The difference between MCV1 and YFV coverage ranges from -5 percentage points (ie. MCV1 is lower than YFV) to +4 percentage points (MCV1 is higher than YFV) across admin1 regions over the years 2018-2022.

3a. By service indicator and year, list the top 2 regions (using the regional coverage estimates) that seem to have implausibly high coverage.

BCG

* 2018: Northern and Brong-Ahafo
* 2019: Northern and Ahafo
* 2020: Northern and Ahafo
* 2021: Northern and North East
* 2022: Northern and North East

DTP1

* 2018: Northern and Brong-Ahafo
* 2019: Northern and North East
* 2020: Northern and North East
* 2021: Northern and Ashanti
* 2022: Northern and Ashanti

DTP3

* 2018: Northern and Brong-Ahafo
* 2019: Northern and Savannah
* 2020: Northern and North East
* 2021: Northern and Ashanti
* 2022: Northern and Ashanti

MCV1

* 2018: Northern and Brong-Ahafo
* 2019: Northern and Ahafo
* 2020: Northern and North East
* 2021: Ahafo and Ashanti
* 2022: North east and Ahafo

YFV

* 2018: Northern and Brong-Ahafo
* 2019: Northern and Ahafo
* 2020: Northern and North East
* 2021: Ahafo and Ashanti
* 2022: Northern and North East

3b. List two good justifications for why the seemingly implausible high coverage estimates may be correct.

1. Migration – There may be cases where children live near the border of regions, but they may cross into the next region to get vaccinated as the health facility could be closer to their home despite being in the next region. As such, these children would be counted in the denominator of their home subregion and the numerator of a neighboring region. Therefore, coverage in the region they were vaccinated in could potentially be greater than 100% as a result of numerator-denominator mismatch.
2. Poor denominators – The number of doses (numerator) may be correct, however, there may be an issue with the denominator being too low. Denominators are often based on census data which are updated each year using estimated growth rates. Distortions grow over time when using this method, even if the census was originally of good quality. Implausibly high coverage can happen when the estimated denominator from a old census is lower than the actual number of children. This could be overcome by using facility-based denominators.

3c. List three good reasons why the seemingly implausible coverage estimates may be wrong

1. Numerator data entry errors – The recorded number of (or tallying of) doses administered could be entered incorrectly either accidentally or deliberately. This could be overcome by optimal design or paper or electronic records, training and regular supervision of data entry staff.
2. Numerator contamination – Older children who receive vaccine doses through catch-up activities or supplementary immunization activities (SIAs) may be included in the numerator. The denominator for DTP3 for example is number of surviving infants to age 1 year. Therefore, older children would not be included in the denominator, but would be included in the numerator. This could be overcome by taking steps to record late vaccination administration.
3. Incorrect recording of doses – There may be instances where the first dose of DTP vaccine is incorrectly recorded as the third dose of DTP. This could lead to implausibly high DTP3 for example. This could be overcome by training and validating against home-based records.