## Name\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

# You have 45 minutes to answer the question below and do the R (or STATA) exercise - This is the R version of the test; if you prefer to use STATA, please open the other Word Document with prefix STATA.

## Part 1

How would you respond to the following email?

Dear HIV Section

I have been asked to provide a very brief summary on whether HIV incidence is decreasing in Blackpool. Please can I have this ASAP?

Head of Blood borne viruses and STIs

Background information (fake data for this assessment)

In 2021, Blackpool started participating in an opt out HIV testing programme in emergency department settings

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 |
| New HIV diagnoses | 25 | 13 | 24 | 14 | 6 | 16 | 28 |
| *Of which diagnosed as part of ED opt out testing* | NA | NA | NA | NA | NA | 10 | 15 |
| Number of HIV tests in sexual health services | 2100 | 2245 | 2670 | 2985 | 3200 | 2269 | 2700 |
| Number of tests in ED settings | NA | NA | NA | NA | NA | 907 | 1123 |

Your response:

Dear Head of Blood borne viruses and STIs,

Based on the available data, there has been variation in the incidence of HIV in Blackpool, as indicated by the sexual health testing data. However, this data does not provide a complete picture of HIV incidence in Blackpool, as it does not capture the entire population.

From the data, there was a significant decrease in the detection rate in 2020, followed by an increase in subsequent years. This trend could be attributed to the implementation of increased testing in emergency departments (EDs) through the opt-out HIV testing programme. The rise in detection rates after 2020 suggests that the ED opt-out testing may have played a role in increasing the number of new HIV diagnoses.. See the table below

|  |  |  |  |
| --- | --- | --- | --- |
| Year | New Diagnoses | Number Tested | Detection Rate (%) |
| 2016 | 2100 | 25 | 1.19 |
| 2017 | 2245 | 13 | 0.58 |
| 2018 | 2670 | 24 | 0.90 |
| 2019 | 2985 | 14 | 0.47 |
| 2020 | 3200 | 6 | 0.19 |
| 2021 | 2689 | 16 | 0.60 |
| 2022 | 2700 | 28 | 1.04 |

The detection rate hit its lowest point in 2020 at 0.19%, likely due to reduced testing during the COVID-19 pandemic. By 2022, the detection rate increased to 1.04%, largely driven by expanded testing efforts, particularly in EDs through the opt-out programme. This suggests that targeted testing strategies, such as the ED opt-out programme, are effective in identifying new cases and improving detection rates.

More population data will need to be able to determine accurately the HIV incidence in black pool.

Thanks.

Aneke Chukwunonye

Senior HIV/STI Surveillance and Prevention Scientist

## Part 2 -**R exercise**

Here is a small section of a R script

df\_2022=df\_2022%>%

group\_by(harspid,reportingyear)%>%

mutate(tot\_att\_year=n())

df\_2022=df\_2022%>%

arrange(attendancedate)%>%

mutate(att\_order\_year=1:n())%>%

ungroup()

for(i in 2012:2015){

df\_2022=df\_2022%>%

mutate(!!(paste0("lastattendancetag\_",i)):=

ifelse(reportingyear==i&tot\_att\_year==att\_order\_year,1,0))

}

df\_2022=df\_2022%>%

arrange(harspid,attendancedate)%>%

mutate(period2nextattendance=ifelse(lead(harspid)==harspid,

lead(attendancedate)-attendancedate,NA))

df\_2022=df\_2022%>%

mutate(period2nextattendance=ifelse(is.na(period2nextattendance)&reportingyear!=2021,

as.Date("2021-01-01")-attendancedate,period2nextattendance))

summary(df\_2022$period2nextattendance)

df\_2022=df\_2022%>%

mutate(tag\_more15m=ifelse(period2nextattendance>456&!is.na(period2nextattendance),

1,0))

table(df\_2022$tag\_more15m)

**Data dictionary**

harspid – unique patient identifier in HARS database

attendancedate – date of attendance

reportingyear – patient’s year of attendance – patients can attend several times in the same year and in separate years

**Questions**

1. Could you please explain the difference between lines 1:3, and 5:8

Lines 1-3 count the total number of attendances per year for each patient.

Lines 5-8 order each attendance within the year to track the order in which attendances occurred.

1. What does the loop on lines 10:14 do?

The loop creates a new column for each year between 2012 and 2015. Each column is named "lastattendancetag" followed by the year. For each year, it checks if the reporting year matches the current year and if the attendance is the last one for that patient in that year. If both conditions are true, it assigns a value of 1 (indicating the last attendance for that year); otherwise, it assigns 0.

1. Could you change line 10 so the loop goes over all values of reportingyear?

for(i in unique(df\_2022$reportingyear)){

df\_2022 = df\_2022 %>%

mutate(!!(paste0("lastattendancetag\_", i)) :=

ifelse(reportingyear == i & tot\_att\_year == att\_order\_year, 1, 0))

}

1. On line 23, what does the function as.Date() do?

The as.Date() function converts a string representing a date (in this case, "2021-01-01") into a date object that R can use for date-based calculations, comparisons, and operations.

1. On line 28, what would be the reason for adding !is.na(period2nextattendance)?

The condition !is.na(period2nextattendance) ensures that only non-missing values of period2nextattendance are included in the calculation. This prevents the function from assigning TRUE to missing (NA) values, ensuring the tagging process only considers valid data and avoids errors caused by missing value