# General Information

The list of chosen columns; It can be changed inside the program.

|  |  |
| --- | --- |
| Sheet | Selected Columns |
| OutputData | E, J, Q, T, V, W, X, DX, [OS-SV] |
| AllVisit | E, H, I, J, L, N, O, P, Q, [R-V], AA, AB, [AC-AF], AH, [CQ-EQ], [LZ-MG], [MT-MX], [MY-NH] |
| DiagnosisData | D, E |
| MedicationData | D, G, J, Q |

# Combining OutputData and AllVisit

* **First\_Biologic (StartDate)** < **AV\_Visit\_Date**
  + All samples of this patient in **AllVisist** sheet will be ignored. Patient with PatientID=9 is an example of this case, as shown below:

|  |  |  |
| --- | --- | --- |
| First\_Biologic(StartDate) (or DX) | AV\_Visit\_Date (or H) | Note |
| 01/09/2009 | 19/01/2021 | Since the Biological treatment date is before all other date, all rows from **AllVisist** will be ignored. Meaning for PatientID=9 columns of **AllVisist sheet** was NaN. |
| 09/04/2021 |
| 20/01/2022 |
| 14/04/2022 |
| 28/07/2022 |
| 27/10/2022 |
| 19/04/2023 |

* **First\_Biologic (StartDate)** Does not exist.
  + It returns the row with the latest date for this patient in **AllVisist** sheet. Patient with PatientID=11 is an example of this case, as shown below:

|  |  |  |
| --- | --- | --- |
| First\_Biologic(StartDate) (or DX) | AV\_Visit\_Date (or H) | Note |
| Not exist. | 13/01/2021 | Since there is no biological treatment date, last row (18/08/2023) was chosen. |
| 18/08/2023 |

* **Remaining cases**
  + Handled as mentioned in our meeting. Patient with PatientID=4 is an example of this case, as shown below:

|  |  |  |
| --- | --- | --- |
| First\_Biologic(StartDate) (or DX) | AV\_Visit\_Date (or H) | Note |
| 24/09/2020 | 14/08/2019 | The columns of row with H=**17/08/2020** waschosen. |
| **17/08/2020** |
| 24/09/2020 |
| 17/12/2020 |
| 10/03/2021 |
| 02/06/2021 |
| 23/03/2022 |
| 15/06/2022 |
| 07/09/2022 |

# Combining OutputData and DiagnosisData

The new column named **hasRheumatoidArthritis** was added to **OutputData sheet** with value [0-3]

* 571 patients do not have Rheumatoid Arthritis. Their values in **hasRheumatoidArthritis** are **zero**.
* 505 patients have one row in DiagnosisData. Their values in **hasRheumatoidArthritis** are **one**.
* 6 patients have two rows in DiagnosisData. Their values in **hasRheumatoidArthritis** are **two**.



* 1 patient has three rows in DiagnosisData. Its value in **hasRheumatoidArthritis** is **three**.



# Combining OutputData and MedicationData

Newly added columns

|  |  |
| --- | --- |
| Column Name | Descriptions |
| Multi | There are same cases when patient took the medication in multiple periods. For example, between 2019 to 2020; then stop and started again between 2021 to 2022. This column how for how many times the medication was taken. The default value is 1.   * **Multi == 0** 🡪 Have never taken this medication * **Multi == 1** 🡪 Take this medication in one period * **Multi > 1** 🡪 Take this medication in multiple period |
| Period | For how long the patient took the medicant. It is in day format; for example, 30 days. There are some special cases:   * If **Multi > 1** 🡪 Sums all periods. * If **isActive == True and** **Multi == 1** 🡪 the value is **-1** * If **isActive == True and** **Multi > 1** 🡪 Sums the period of previous sessions |
| isActive | Is a patient still use this medication? **True** for yes, otherwise return **False** |

The final output for column names of this sheet will be (**MedicamentGroupName** + **ColumnName**) As shown below:

A white rectangular object with black text

Description automatically generated