Dataset generation for M&S

QC Checklist

**1. Dataset generation strategy**

A total of three files were provided (S24008\_Final Resul\_보고서용\_240527.xlsx, G001\_Ocular PK 분석결과\_(16주차)*2024.06.24 (공유본).xlsx, and G001\_Ocular PK 분석결과*(16주차)\_2024.06.24 (공유본).xlsx). To facilitate data tidying in R, all xlsx files were converted to csv format before importing them into R environment. The contents of the three files were Aflibercept concentrations (ng/mL) over time for various dose groups(G1~G11), with a total of 7 sites (in order: aqueous humor, vitreous, iris, retina, choroid, optic nerve, and serum) and concentration values divided into left and right eye. For ID, IDs specified in excel files were used as is. TIME was divided into time (TIME), day (DAY), and week (WEEK). For DV, concentration values that were BQL were all treated as 0 (DV=0, MDV=1). LEFT column (1: left, 0: right) was created to distinguish between left and right eye, for the SITE column, each site was represented numerically in the order mentioned above. For convenience, GROUP (G1 ~ G11), BLQ (5ng/mL), DRUG (4: Vehicle, 3: 3TF, 2: Aflibercept, 1: AIO), and ROW\_Number columns were added. After tidying, the dataset was arranged by ID, and TIME. The dataset generated using R was evaluated with the original dataset to identify the possible errors.

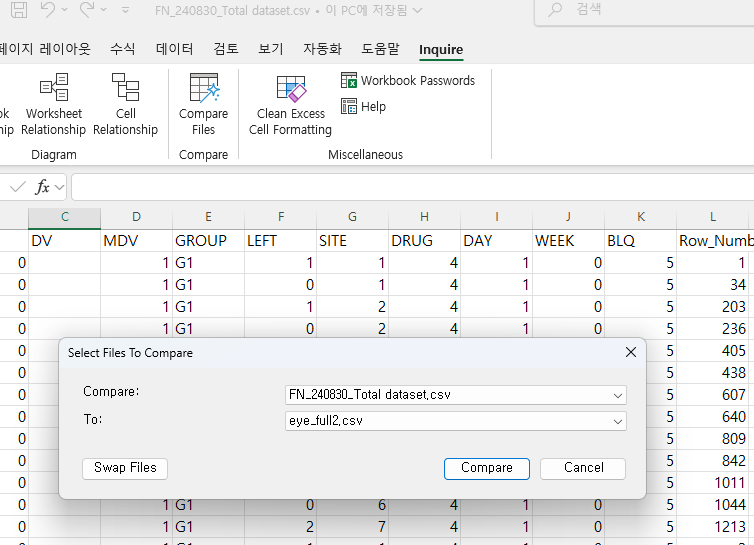
**2. Result**

Dataset generation was conducted using R (Version 4.3.1) and the `tidyverse` package (version 2.0.0). The final dataset consisted of 12 columns and 1547 rows including 1106 rows with concentration and 441 rows without concentration (BQL). For the 16-week and 24-week data, 1000 was added to the existing IDs to avoid overlap. The remaining columns were created as specified above, and the dataset was arranged by ID, and TIME to create the final dataset. The R code used to generate the dataset is attached below as a .txt file (QC\_NONMEM.txt, Supplementary 1).

**2. Dataset evaluation**

Dataset comparison was performed using the `inquire` Addin provided in Excel (Figures). `Excel inquire` is an additional feature provided by Excel used for comparing different Excel files. Comparing two different datasets revealed that the number of rows and columns were all observed to be identical but `inquire` reported 1107 unidentical rows (Inquire Report, Supplementary 2). First, the concentration collected location was named differently as CMT and SITE, which was corrected to SITE. Secondly, in the DV column, 1106 rows were reported as different. This discrepancy appeared to be from variations in decimal point process by R, causing nominally identical values to be recognized as different due to slight difference in decimal point. As the data generated by the above R code showed all decimal points as single digits, it appeared that errors related to decimal points occurred during the process of converting raw data to .csv and R tidying. To verify this, the original dataset was filtered using R to include only rows with concentration values, resulting in 1106 rows, which confirms the number of rows reported as different by the Excel inquire. No other notable differences were observed.

**2. Figures**



텍스트, 스크린샷, 소프트웨어, 도표이(가) 표시된 사진

자동 생성된 설명

Excel Addin `Inquire`