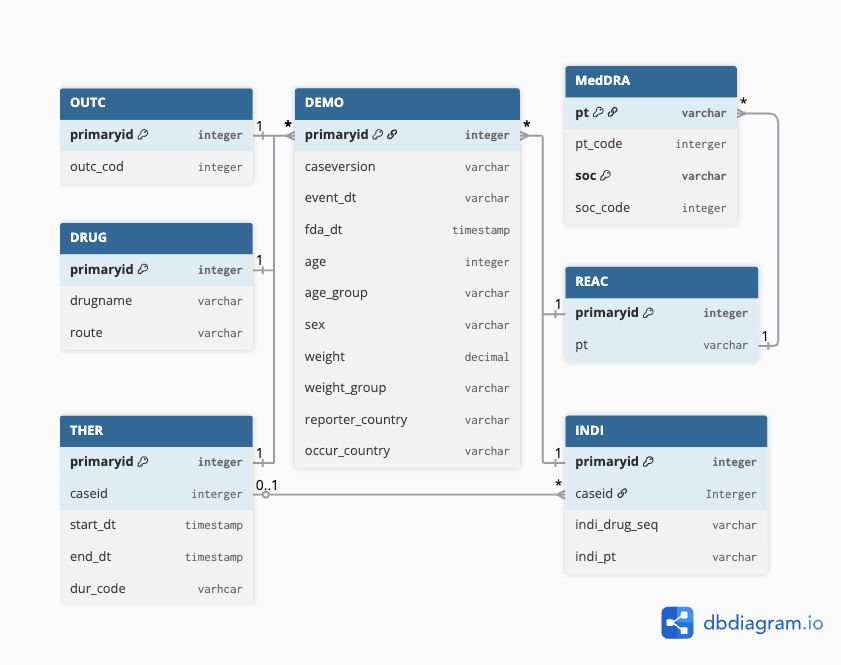
**Supplementary Information**

**FAERS Quarterly Data Extract files (ASCII)**

|  |  |  |
| --- | --- | --- |
| **Abbreviations** | **File Name** | **Content** |
| DEMO | DEMOyyQq.TXT | DEMOGRAPHIC |
| DRUG | DRUGyyQq.TXT | DRUG |
| REAC | REACyyQq.TXT | REACTION |
| OUTC | OUTCyyQq.TXT | OUTCOME |
| RERP | RERPyyQq.TXT | REPORT SOURCE |
| THER | THERyyQq.TXT | THERAPY |
| INDI | INDIyyQq.TXT | INDICATIONS |

**FAERS Dataset Relationship**

****

**Abbreviations Glossary**

***95% CI:*** 95% Confidence Interval

***ADR:*** Adverse Drug Reaction

***AE:*** Adverse Event

***AER:*** Adverse Event Report

***AUROC:*** Area Under the Receiver Operating Characteristic

***BCPNN:*** Bayesian Confidence Propagation Neural Network

***DSA***: Disproportionality Statistical Analysis

***FAERS:*** FDA Adverse Event Reporting System

***GB:*** Gradient Boosting

***IC025:*** LowerBound of Important Component

***KAERS:*** Korea Adverse Event Reporting System

***MedDRA:*** Medical Dictionary for Regulatory Activities

***MGPS***: Multi-item Gamma Poisson Shrinker

***MHRA:*** Medicines and Healthcare products Regulatory Agency

***PRR***: Proportional Reporting Ratio

***PT:*** Preferred Term

***RF:*** Random Forest

***ROR:*** Reporting Odds Ratio

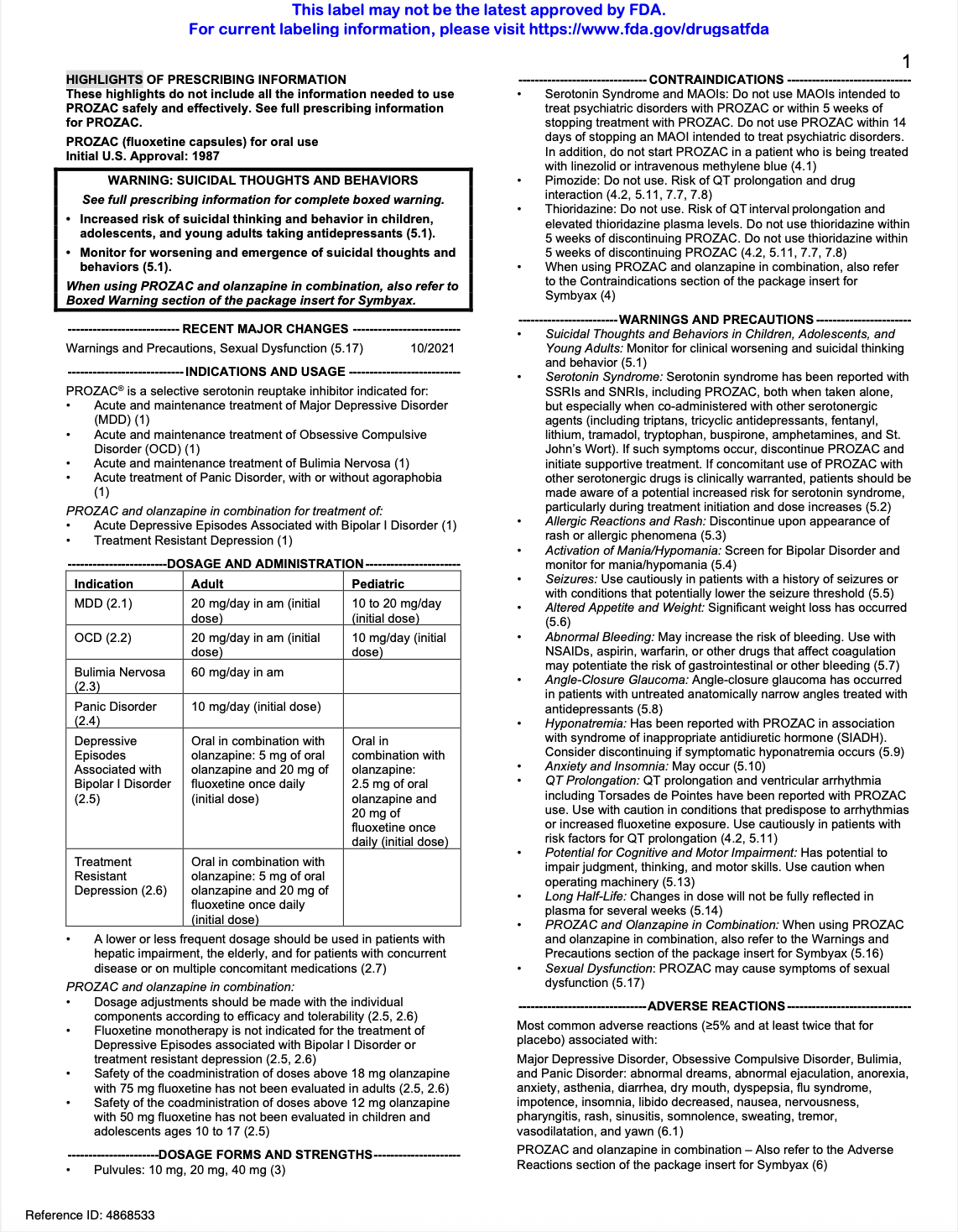
***SHAP***: SHapley Additive exPlanation

***SMOTE:*** Synthetic Minority Over Sampling Technique

***SOC:*** System Organ Class

***SSIR:*** Selective Serotonin Reuptake Inhibitors

## **Fluoxetine FDA Label Change Record (06/10/2021) – Warnings and Precautions**



**Data Preprocessing Details**

The changes in file formats and filed structures necessitated a phased approach to data concatenation. Accordingly, the concatenation process is divided into three distinct phases. The first phase spans from 2004Q1 to 2012Q3 during which FDA was still using Individual Safety Report system (ISR) to record the adverse event reports. The second phase is the transition period from 2012Q4 to 2014Q2, where they upgraded to the CASEVERSION system. The third phase covering 2014Q3 to 2024Q4 which substantially revised and updated several new indicators in the datasets.

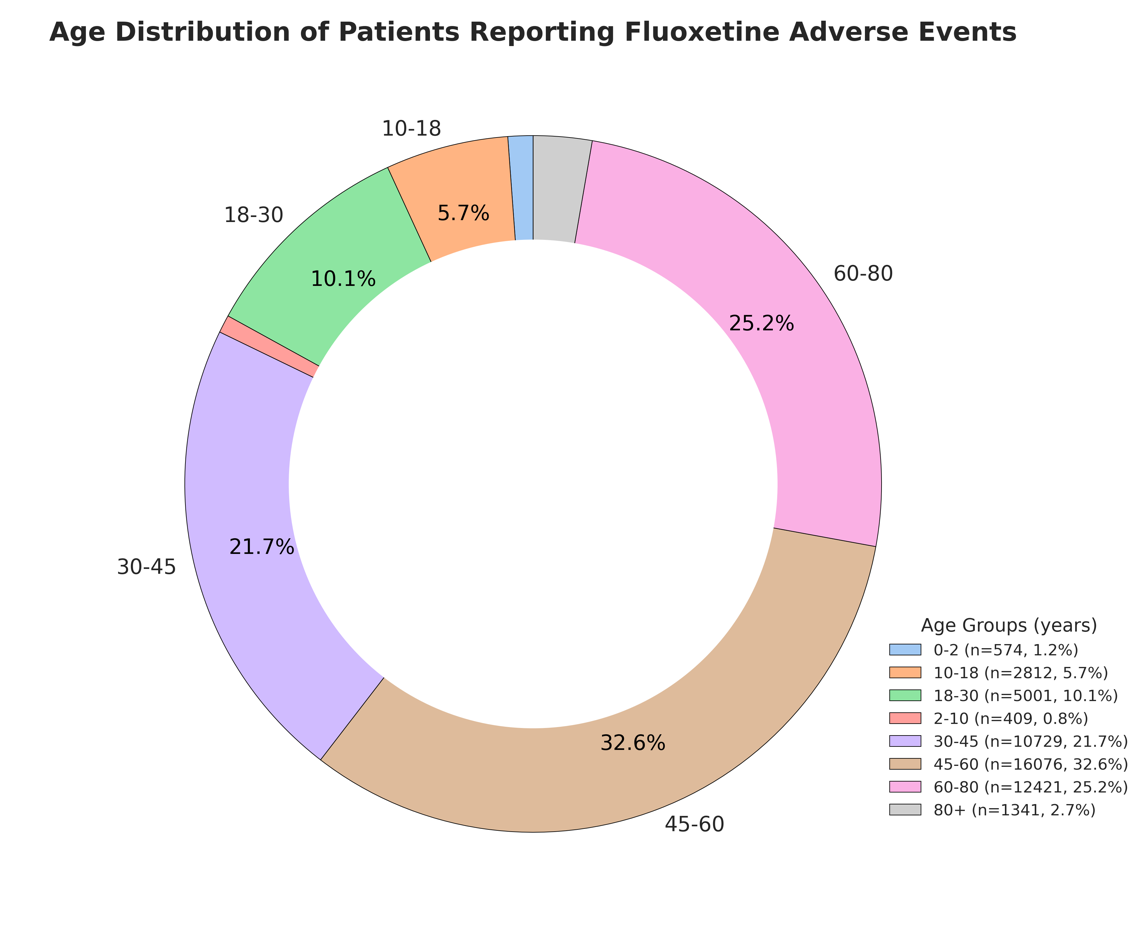
Subsequently, for each set of quarterly files, the header row from the first file was preserved while automatically removing headers from subsequent files. Batch processing across all 84 independent quarterly files improves cost-effectiveness and maintained the temporal integrity of the data. Eventually, three merged files were created, one of reach phase, across all seven FAERS components (see Appendix).

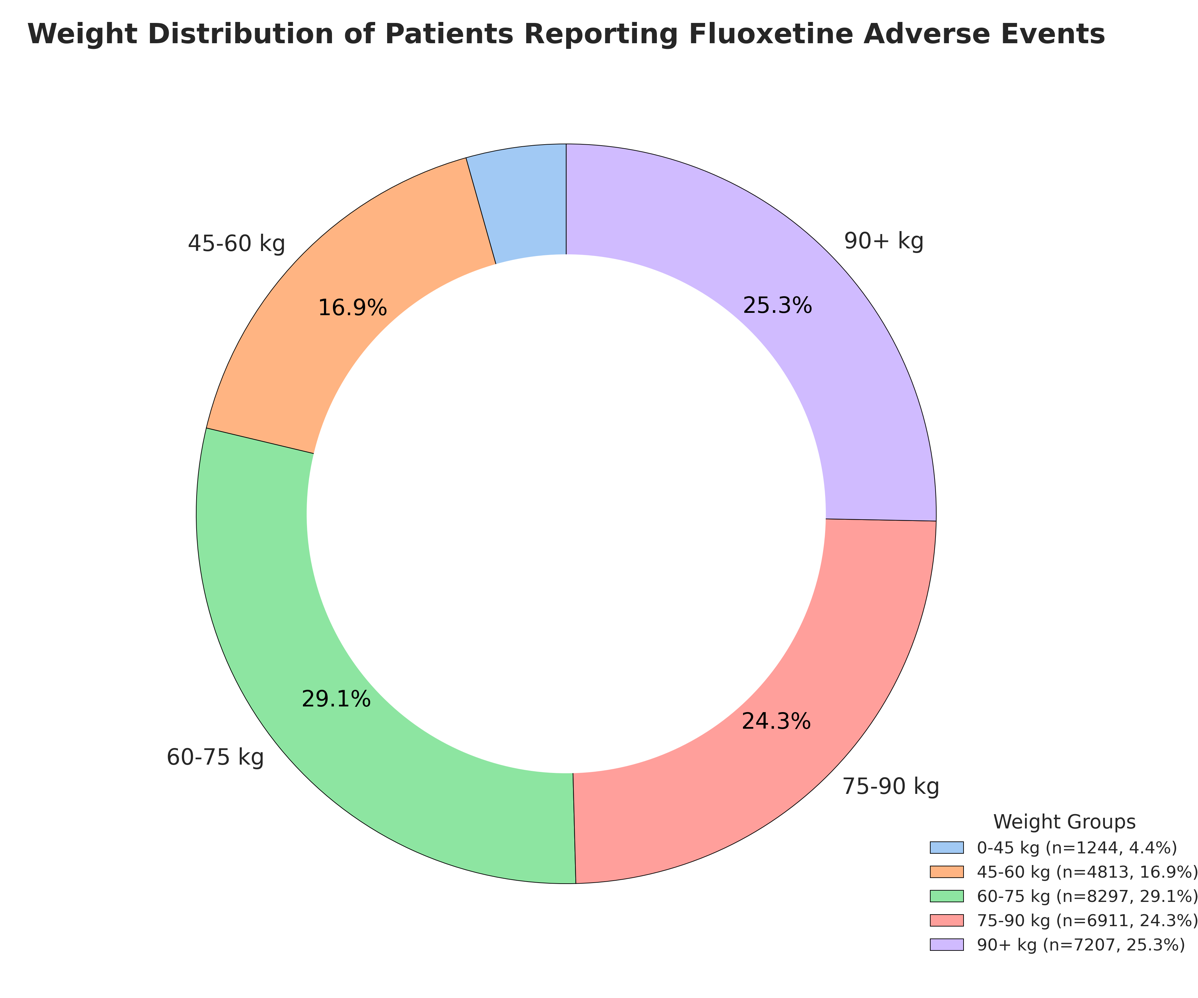
Due to inconsistencies naming convention in early file formats, such as use of 'ISR' instead of 'primaryid' and 'OUTC\_COD' instead of 'outc\_code', column names therefore were standardized across all versions before concatenation to ensure consistency. After standardization, only the essential columns were retained for analysis, specifically the case identifier ('primaryid') and the adverse reaction term ('pt'). This selective retention of relevant columns substantially reduced data complexity while preserving analytical value.

The medication data required extensive preprocessing owing to its structural complexity. Particularly, the DRUG dataset was enriched by integrating timeline information from THER dataset, so that allows to calculate medication exposure duration. To handle inconsistent recording of duration units, a mapping scheme was developed to convert various time units into standardised unit of days. These duration values were then categorized into clinically meaningful intervals to facilitate stratified analysis. Similar standardisation was also applied to medication routes of administration, reduced over 199 route descriptions to 106 consolidated categories, thereby improving analytical tractability.

Demographic dataset presented a unique challenge due to the sensitive and variable natural of personal information. Age and weight variables were standardized and categorised into clinically relevant groups. We further reconciled country codes and names recorded by mapping them with two-character country codes based on ISO.

**Fluoxetine Adverse Event Extended Visualizations**





A graph of different colored bars

AI-generated content may be incorrect.A graph showing a number of events

AI-generated content may be incorrect.