

Original C1 Dataset
(n = 163,146;
patients: 85,722)

Original Prosecutors Office
(n = 568,297;
Causes= 494,929;
Rel.=568,290;
RUC_Vic_Imp=541,595;
individuals= 74,833)

Data wrangling and normalization process

- Remove duplicated entries
- Overlapping treatments of patients
- Intermediate treatment events (continuous referrals)

Data wrangling and normalization process

- Filter crimes committed after study follow-up period(p= 1,206; RUCs= 1,308;n= 1,413)
- Remove duplicated entries(p= 7; RUCs= 7;n= 7)
- Correct dates (birth, comission of crime, end of judicial proceedings), missing nationality and sex
- Erase entries with missing values in comission of crime, end of judicial proceedings(p= 13; RUCs= 57;n= 61)
- Erase entries with values in comission of crime when minor to 14 years old after imputation(p= 557; RUCs= 895;n= 557)
- Filter crimes committed before study follow-up(p= 17,506; RUCs= 22,776;n= 28,371)
- Filter records with cause of end of the proceedings= administrative annulment(p= 1,053; RUCs= 1,163;n= 1,503)
- Filter records with cause of end of the proceedings= grouped to another case(p= 15,739; RUCs= 24,861;n= 28,024)

C1 Dataset
(n = 109,756;
patients: 85,048)

O.P. Dataset
(n= 508,236;
individuals= 74,535)

27
56

- Discard observations coded as victims rather than ofenders (p= 56,301; RUCs= 155,538;n= 171,203)
- Discard observations depending on the values of the end of the proceedings (p= 68,268; RUCs= 279,609;n= 311,147)
- Discard observations without birth date (p= 7; RUCs= 1(had no case ID);n= 7)
- Discard observations coded as referrals, deaths, censored at baseline treatment or with ongoing treatments (p= 18,411; RUCs= 4,351;n= 20,160)
- Long-to-wide relationships/crimes, end of judicial proceedings, penalty
- Group crimes into violent, drug-related, etc.

Merged database