### Section 1. Specifics on the data wrangling

We started with a cohort of SENDA patients used in other publications (1, 2). Then, we managed to link these patients with Prosecutor’s Office data regarding offences where patients were involved (as victims or offenders).

However, we focused on cases in which patients were found as offenders. We did not require any minimum duration of first SUT enrolment for study inclusion to reduce the risk of selection bias and strengthen the generalizability of results. Readers must know that dropouts occur due to voluntary or involuntary causes. Voluntary causes include when the patient explicitly states their will to discontinue treatment or non-attendance and/or lose contact with the patient for a period equal to or greater than 45 days. And involuntary causes include interruptions due to serious misconduct against treatment norms. Also, we excluded external referrals.

The time to event was defined after treatment outcome (completion or dropout). Thus, we avoided survival bias due to a dropout explained by the event (reverse causation).

The process is summarised in Supplementary Table 1.

Supplementary Table 1 Sample flow chart of main methodological decision per each database

|  |  |  |  |
| --- | --- | --- | --- |
| **Order** | **Description** | **Total patients** | **% Changed from the previous step** |
| **Database 1: Patients in substance use treatment** | | | |
| 1 | Individuals with ongoing SUT between 2010 and 2019 | 85,048 | 100% |
| **Database 2: Prosecutor's Office information** | | | |
| 1 | Patients of Database 1 that had records in the Prosecutor’s Office | 74,833 | 100% |
| 2 | Excluded records with dates of commission after November 13th, 2019 (n= 47) | 74,786 | 99.90% |
| 3 | Erased records with missing information, discrepancies in ages between SENDA, and aberrant ages of commission (n= 41) | 74,745 | 99.90% |
| 4 | Excluded records with at least one of the conditions: administrative annulment, grouped to another criminal case, commission dates earlier than January 01, 2010, and duplicated records (n=210) | 74,535 | 99.70% |
| 5 | Excluded records of patients with at least one of the following conditions: found as a victim (n=20,624), if the patient did not receive a sentence (n=23,667) | 49,970 | 67.04% |
| **Database 3: Joined database (Database 1 + Database 2)** | | | |
| 1 | Individuals with ongoing SUT between 2010 and 2019 | 85,048 | 100.00% |
| 2 | Excluded individuals referred by other prior SUT (n=8,657) | 76,391 | 89.82% |
| 3 | Excluded individuals with ongoing treatments at the date of information retrieval (November 13th, 2019) (n=5,521) | 70,870 | 92.77% |
| 4 | Excluded individuals with missing dates of birth (n=7) | 70,863 | 99.99% |
| 5 | Excluded patients without treatment outcomes (n=9) | 70,854 | 99.99% |

> nrow(Base\_fiscalia\_v15f\_grant\_23\_24\_long2)

[1] 90075

> length(unique(Base\_fiscalia\_v15f\_grant\_23\_24\_long2$hash\_key))

[1] 72404