Original C1 Dataset

(

n = 163,146;

users: 85,722)

Data wrangling and normalization process

C1 Dataset

n = 109,756;

(

users: 85,048)

Sample selection

•Discard men (users=63,625)

•Discard posterior treatments (>2) (n= 6,944; users=4,853)

•Discard population under 18 years of age (users=114)

C1 Dataset

(

n = 21,378;

users: 21,378)

Impute missing values: (n=1,587)

•Duplicated entries

•Intermediate events of treatment (continuous referrals)

Imputed C1 Dataset

: 21,378)

(

users

Result of the imputation of missing values

•Logically inconsistent candidates for imputation

•Ties in candidates for imputation