

创建队列代码分析

Code block

```
1
2 > str(ost)
3 Classes 'data.table' and 'data.frame':      6163698 obs. of  5 variables:
4  $ IAIN                      : num  24122281 24122281 24097014 24114801 24114801
5  $ patient_date_of_birth      : Date, format: "1974-02-28" "1974-02-28" "1978-06-
6  $ patient_sex                : hvn_lbl [1:6163698] 1, 1, 1, 1, 1, 1, 1, 1, 2,
7  ..@ labels: Named num 1 2
8  .. ..- attr(*, "names")= chr [1:2] "Male" "Female"
9  $ paid_month_end_date        : Date, format: "2009-04-30" "2009-04-30" "2009-04-
10 $ patient_health_board_name: chr  "NHS GRAMPIAN" "NHS GRAMPIAN" "NHS GRAMPIAN"
11 - attr(*, ".internal.selfref")=<externalptr>
12
13
```

旧代码

1. 初始设置与数据加载

Code block

```
1 # Libraries -----
2 rm(list=ls())
3 library(tidyverse) # 数据清洗和分析
4 library(tidylog) # 提供数据操作日志
5 library(lubridate) # 日期处理
6 library(data.table) # 高效数据处理
7 library(janitor) # 数据清理
8 library(retroharmonize) # 数据标准化
9 library(phsmethods) # PHS特定方法
```

```
10 gc() # 垃圾回收释放内存
11
12 # 加载OST处方数据
13 orts <- read_rds("/OST_prescriptions.rds")
14 range(orts$PaidDate) # 检查日期范围
```

功能：

- 清理环境并加载必要的R包
- 读取OST处方数据并检查日期范围
- 关键包说明：`data.table` 用于大数据处理，`lubridate` 处理日期，`tidylog` 记录数据操作

2. 参数定义

Code block

```
1 my_origin_date <- as.Date("2000-01-01") # 基准日期
2 admin_start <- as.Date("2010-01-01") # 研究开始日期
3 admin_end <- as.Date("2024-12-31") # 研究结束日期
4 gap_length <- 62L # 处方间隔阈值(天)
5 censor_length <- 1826L # 审查期(5年)
6 presc_length <- 48L # 处方有效期(天)
```

关键参数：

- `gap_length=62`：超过此天数视为新治疗周期
- `presc_length=48`：处方有效期为48天
- `censor_length=1826`：5年随访期

3. 数据预处理

```

1  orts2 <- orts %>%
2    mutate(across(c("IAIN"), as.character)) %>%
3    distinct(IAIN, PaidDate, .keep_all=TRUE) %>% # 去重
4    arrange(IAIN, PaidDate) %>%
5    mutate(
6      presc_start = as.Date(PaidDate - days(60)), # 处方开始日期(提前60天)
7      presc_end = as.Date(presc_start + days(presc_length)) # 处方结束日期
8    )

```

处理逻辑：

1. 确保患者ID为字符型
2. 按患者和支付日期去重
3. 定义每个处方的有效期间隔（支付日前60天至后48天）

4. 治疗周期计算

Code block

```

1  prescriptions <- orts2 %>% setDT()
2  prescriptions[, presc_between := difftime(presc_start, shift(presc_end,
3    type="lag"), units="days"), by=IAIN]
4  prescriptions[, presc_between := fifelse(is.na(presc_between), 0L,
5    presc_between)]
6  prescriptions[, gap := fifelse(presc_between > gap_length, 1L, 0)] # 间隔标记
7  prescriptions[, episode := rleid(gap), by=IAIN] # 治疗周期编号

```

核心算法：

- 计算相邻处方的间隔天数
- 超过62天的间隔标记为新治疗周期
- 使用 `rleid` 生成连续的治疗周期编号

5. 治疗周期汇总

Code block

```
1 episodes_prescriptions <- prescriptions[, .(
2   ep_start = min(presc_start),
3   ep_end = max(presc_end)
4 ), by = .(IAIN, episode)]
```

输出：

每个患者的每个治疗周期的开始和结束日期

6. 队列时间定义

Code block

```
1 cohort_dates <- episodes_prescriptions[, .(
2   date_of_entry = min(ep_start), # 首次处方日期
3   date_of_exit = max(ep_end) + censor_length # 最后处方日期+5年
4 ), by = IAIN]
5
6 # 财务年度调整
7 cohort_dates[, FY := phsmethods::extract_fin_year(date_of_exit)]
8 cohort_dates[, date_of_exit := case_when(...)] # 按财务年度截断
9
10 # 合并死亡日期
11 cohort_dates <- merge(cohort_dates, patient_summary[, .(IAIN, date_of_death)],
12   by="IAIN", all.x=TRUE)
12 cohort_dates[, date_of_exit := pmin(date_of_exit, date_of_death, na.rm=TRUE)]
```

关键处理：

- 定义观察期：从首次处方到最后处方+5年
- 调整日期匹配财务年度(4月1日至次年3月31日)
- 死亡日期早于观察期结束时，以死亡日期为准

7. 每日记录展开

Code block

```
1 dt <- cohort_dates[, .(day = seq(from=date_of_entry, to=date_of_exit,  
  by="day")), by=IAIN]
```

结果：

为每个患者生成每日记录的时间序列数据

8. 治疗状态标记

Code block

```
1 # 标记在治疗期间的日期  
2 episodes_prescriptions[, .(day=seq(from=ep_start, to=ep_end, by="day")), by=.  
  (IAIN, episode)] %>%  
3   merge(dt, by=c("IAIN", "day"), all.x=TRUE)  
4  
5 dt[, status_oat := fifelse(!is.na(episode), 1L, 0L)] # 治疗状态标记
```

逻辑：

- 将治疗周期展开为每日记录
 - 合并到主数据并标记治疗状态(1=在治疗，0=未治疗)
-

9. 协变量添加

```

1 # 添加人口统计学变量
2 dt <- merge(dt, patient_summary[, .(IAIN, DOB, Gender)], by="IAIN")
3 dt[, age := floor(as.numeric(day - DOB)/365.25)] # 计算年龄
4 dt[, sex := fifelse(Gender==1, "M", "F")]
5
6 # 添加住院(过量)记录
7 dt <- merge(dt, nfod[, .(IAIN, day, nfod=1L)], by=c("IAIN","day"), all.x=TRUE)
8 dt[, nfod := fifelse(is.na(nfod), 0L, 1L)]

```

新增变量：

- 年龄、性别
- 非致命过量(nfod)标记

10. 最终数据处理

Code block

```

1 # 筛选研究期间记录
2 dt <- dt[day %between% c(admin_start, admin_end)]
3
4 # 迁移数据处理
5 dt <- merge(dt, migration_lookup, by="IAIN", all.x=TRUE)
6 dt[, migration_flag := fifelse(is.na(migration_flag), 0L, 1L)]
7
8 # 退出标记(死亡或迁移)
9 dt[, exit_flag := fifelse(acm_flag==1 | migr_flag==1, 1L, 0L)]

```

质量控制：

- 限制在2010-2024研究期间
- 处理迁移数据(移除迁移期间记录但保留迁移事件标记)
- 定义观察终点(死亡或迁移)

11. 分析数据集生成

Code block

```
1 episodes <- dt[, .(  
2   int_start = min(day),  
3   int_end = max(day),  
4   days_at_risk = .N,  
5   nfod_count = sum(nfod)  
6   ), by=.(IAIN, period, age_group, sex, FY)]
```

最终输出：

- 每个患者的治疗周期汇总数据
- 包含风险天数、过量次数等关键指标

数据集及其作用

1. 原始OST处方数据 (`**orts**`)

文件： `cohort_OST_prescriptions.rds`

关键字段：

- `IAIN`：患者唯一标识符
- `PaidDate`：处方支付日期
- `patient_date_of_birth`：患者出生日期
- `patient_sex`：性别（1=男，2=女）
- `patient_health_board_name`：所属NHS健康委员会

作用：

- 核心分析对象，记录所有阿片类药物替代治疗(OST)处方事件

- 提供治疗暴露（exposure）的时间点和患者基础信息
 - **输出：**生成治疗周期定义和患者观察时间窗的基础
-

2. 患者汇总数据 (`**patient_summary**`)

文件: `patient_summary_acm.rds`

关键字段：

- `IAIN`：患者ID
- `date_of_death`：死亡日期
- `DOB`：出生日期（与OST数据重复，用于校验）

作用：

- 提供患者死亡日期，用于定义观察终点（删失）
 - 验证人口统计学信息的完整性
 - **关键用途：**在 `cohort_dates` 中合并死亡日期以调整观察期
-

3. 非致命过量数据 (`**nfod**`)

文件: `hospitalisations_T.rds`

关键字段：

- `IAIN`：患者ID
- `ADMISSION_DATE`（代码中重命名为 `day`）：住院日期
- 隐含的过量标记（通过字段选择推断）

作用：

- 识别药物相关伤害事件（非致命过量）
- 用于构建结局变量（ `nfod` 和 `nfod_count` ）
- **分析意义：**作为泊松回归的响应变量或竞争风险事件

4. 死亡记录数据 (`**allHMMBs**` `**/**` `**allACM**`)

文件：

- `allHMMBs.rds` : 药物相关死亡
- `allACMs.rds` : 全因死亡

关键字段：

- `IAIN` : 患者ID
- `day` : 死亡日期
- `hmmb_flag` / `acm_flag` : 死亡类型标记

作用：

- `allHMMBs` : 专门捕获药物过量导致的死亡（高优先级结局）
 - `allACM` : 全因死亡用于综合生存分析
 - **处理逻辑**：死亡日期用于截断观察期（右删失）
-

5. 迁移记录数据 (`**migration_lookup**`)

文件： `lookup_CHI_migration_IAIN.rds`

关键字段：

- `IAIN` : 患者ID
- `DATE_TRANSFER_OUT` : 迁移日期
- `int_start` / `int_end` : 迁移时间区间（代码中生成）

作用：

- 识别患者迁移出研究区域的时间
- **处理方式**：
 - 迁移期间的数据被排除（`migration flag=1` 时过滤）

- 迁移事件本身作为退出标志 (`migr_flag=1`)

6. 中间生成的关键数据集

(1) 处方级数据 (`**orts2**` `**/**` `**prescriptions**`)

字段:

- `presc_start` / `presc_end` : 处方有效期间隔
- `presc_between` : 处方间间隔天数
- `gap` : 是否为新治疗周期 (>62天间隔)

信息:

- 将原始处方记录转化为连续治疗周期的基础
- 定义治疗中断 (`gap`) 的标准

(2) 治疗周期数据 (`**episodes_prescriptions**`)

字段:

- `episode` : 治疗周期编号
- `ep_start` / `ep_end` : 周期起止日期

信息:

- 患者接受OST的连续/间断治疗阶段
- 核心分析单位 (如保留率计算)

(3) 队列时间数据 (`**cohort_dates**`)

字段:

- `date_of_entry` : 首次OST日期
- `date_of_exit` : 观察结束日期 (最后处方+5年或死亡)

- `FY`：财务年度（用于时间分层）

信息：

- 定义每个患者的风险时间（time at risk）
- 控制 immortal time bias 的关键

(4) 每日状态数据 (`**dt**`)

字段：

- `status_oat`：当日是否在治疗（1/0）
- `nfod`：当日是否发生非致命过量
- `age` / `sex`：时变协变量

信息：

- 最细粒度的时间序列数据
- 用于计算累积暴露量和时变协变量

(5) 最终分析数据集 (`**episodes**`)

字段：

- `period`：治疗状态（on/off）
- `days_at_risk`：该阶段风险天数
- `nfod_count`：阶段内过量次数

信息：

- 聚合后的分析单元
- 可直接用于泊松回归或生存分析

数据流逻辑总结

1. 原始数据 → 2. 治疗周期定义 → 3. 观察期划定 → 4. 每日状态展开 → 5. 协变量合并 → 6. 分析数据集聚合

各数据集在分析中的作用

数据集	主要用途	关键贡献
orts	治疗暴露	定义OST开始/结束时间
patient_summary	删失处理	提供死亡日期
nfod	结局变量	非致命过量事件
allHMMBs	结局变量	药物相关死亡
migration_lookup	删失处理	迁移事件识别
episodes	分析单元	聚合的治疗周期数据

表：研究结局与协变量的定义及对应 ICD-10 代码（中文对照）

描述	ICD-10 代码
药物相关死亡	
· 精神活性物质所致精神与行为障碍（不含酒精与烟草）	F11, F12, F13, F14, F15, F16, F19
· 意外中毒及有害物质暴露	X40, X41, X42, X43, X44
· 故意自残	X60, X61, X62, X63, X64

· 药物或生物制品攻击	X85
· 意图不明的中毒	Y11, Y12, Y13, Y14
精神/行为障碍住院	
· 精神分裂症及相关精神病	F20–F29
· 情感障碍（情绪障碍）	F30, F31, F32, F33, F39, F34.0, F34.1, F34.8, F34.9, F43.21, F43.23
· 神经症、应激相关及躯体形式障碍	F40–F48（不含 F43.21, F43.23）
· 成人人格与行为障碍	F60–F69
· 注意缺陷多动障碍（ADHD）	F90
非致命过量住院	
· 药物、药品及生物制品中毒	T40.0–T40.6, T42.4, T42.6–T42.7, T43.6

Description	
Drug-related death	
Mental and behavioural disorder due to psychoactive drug use (excluding alcohol and tobacco)	F11, F12, F13, F14,
Accidental poisoning by and exposure to noxious substances	X40, X41, X42, X43,
Intentional self-harm	X60, X61, X62, X63,
Assault by drugs, medicaments and biological substances	X85
Poisoning by undetermined intent	Y11, Y12, Y13, Y14
Admission for mental and behavioural disorder	
Schizophrenia and psychoses	F20, F21, F22, F23,
Mood [affective] disorders	F30, F31, F32, F33,
Neurotic, stress-related and somatoform disorders	F40, F41, F42, F43 (
Disorders of adult personality and behaviour	F60, F61, F62, F63,
Attention deficit hyperactivity disorder	F90
Admission for non-fatal overdose	
Poisoning by drugs, medicaments and biological substances	T40.0, T40.1, T40.2

