# cohortCapr markdown

Guus @TheHyve

2024-07-17

## Concept counts

Retrieve counts for a concept set

```
## Count occurrences of each concept in data

# Establish connection
con <- connect(connectionDetails)
```

```
## Connecting using PostgreSQL driver
```

```
# Get countOccurrences function
source("./R/countOccurrences.R")

# Get links between tables and fields as input
source("./R/table_linked_to_concept_field.R")
```

```
## Sourced links object
```

```
print(links)
```

```
## $condition_occurrence
## [1] "condition_concept_id"   "condition_source_value"
##
## $death
## [1] "cause_concept_id"   "cause_source_value"
##
## $device_exposure
## [1] "device_concept_id"   "device_source_value"
##
## $drug_exposure
## [1] "drug_concept_id"   "drug_source_value"
##
## $measurement
## [1] "measurement_concept_id"   "measurement_source_value"
##
## $observation
## [1] "observation_concept_id"   "observation_source_value"
##
## $procedure_occurrence
```

```
## [1] "procedure_concept_id"    "procedure_source_value"
##
## $specimen
## [1] "specimen_concept_id"    "specimen_source_value"
##
## $visit_occurrence
## [1] "visit_concept_id"    "visit_source_value"
```

```r
conceptCounts <-
  countOccurrences(
    c(3004249, 317009, 3027946, 4029498, 317510, 5083),
    c("condition_occurrence", "procedure_occurrence", "measurement", "observation"),
    links,
    con,
    connectionConfig$cdm_schema,
    connectionConfig$vocabulary_schema,
    save_path = config_oth$save_path_counts
  )
```

```
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## Warning: Input SQL has already been translated, so not translating again
## This warning is displayed once every 8 hours.
```

```
##   |                                                                         |
```

```
## Executing SQL took 0.0262 secs
```

```
##   |                                                                         |
```

```
## Executing SQL took 0.00623 secs
## 'summarise()' has grouped output by 'concept_id', 'concept_name'. You can override using the '.group
```

```r
knitr::kable(conceptCounts, format = "latex", caption = "countOccurrences results",
             col.names = c("concept_id", "concept_name", "domain", "n_person", "n_record", "n_desc_pers
             align = rep.int("l", 7)) |>
  kableExtra::column_spec(c(1,2,3,4,5,6,7), width = rep.int("6em", 7)) |>
  kableExtra::scroll_box(width = "100%", height = "500px")
```

```
## Warning in table_info$align_vector[column] <-
## unlist(lapply(table_info$align_vector_origin[column], : number of items to
## replace is not a multiple of replacement length
```

Table 1: countOccurrences results

| concept_id | concept_name | domain | n_person | n_record | n_desc_person | n_desc_record |
|---|---|---|---|---|---|---|
| 3004249 | Systolic blood pressure | Measurement | 1002 | 13250 | 1002 | 13250 |
| 317009 | Asthma | Condition | 0 | 0 | 71 | 71 |
| 3027946 | Carbon dioxide [Partial pressure] in Arterial blood | Measurement | 3 | 39 | 3 | 39 |
| 4029498 | Seizure disorder | Condition | 14 | 14 | 15 | 24 |
| 317510 | Leukemia | Condition | 0 | 0 | 1 | 1 |
| 5083 | Telehealth | Visit | 0 | 0 | 0 | 0 |

```
# Disconnect
disconnect(con)
```