Regimen Identification for Truven Database

**Description:** A function to identify drug regimens and cycles for a cohort of patients.

**Usage:**

*regimens(*patient\_id\_table *= "ac\_chemo\_anemia\_breast\_pts",* procedure\_id\_table *= "ac\_chemo\_anemia\_breast\_cancer\_procs",* new\_table\_name *= "ac\_breast")*

**Arguments:**

|  |  |
| --- | --- |
| *data\_source* | Data source, currently Truven or Symphony arguments are accepted. |
| *patient\_id\_table* | String name of table in sandbox\_grwi. That lists the enrolids of the patients of interest. |
| *procedure\_id\_table* | String name of table in sandbox\_grwi. That lists the procedures of interest. |
| *base\_data\_schema* | Schema from which to extract data Ex: “truven\_commercial\_claims\_union” or “sandbox\_rwi\_metrics” |
| *base\_data\_table* | Table from which to extract data Ex: “outpatient\_services” or “xem\_pc\_nps\_lot\_tmp\_03\_psa” |
| *new\_table\_name* | String name of name format of the new tables generated for each of the six steps. |
| *treatment\_identifier* | Column name of treatment observed for regimen in *procedure\_id\_table.* Ex: “proc1” |

**Value:**

Final result will be six tables, produced after each step. The final table with regimens and cycles identified will be in:

sandbox\_grwi.*new\_table\_name* + “\_step6”,

where *new\_table\_name* is defined by the user in the third function argument.

**Use:**

An example of use is the finding regimens and cycles of cancer patients to create a clearer picture of a patient’s treatment cycle in order to identify what cancer treatments and frequency of treatment has a high propensity for anemia.

**Run Time:**

**Details:**

* Regimen – A combination of certain drugs that a patient takes
* Cycle – Iteration of administration of a distinct regimen

Step 1: Get patient Claims

* Get every claim of cohort where procedure for chemotherapy, defined in Stephen’s code, is performed.
* Find start of chemo, svcdate and count of daily procedures.

Step 2: Get regimen 1 potential service dates

* Find 28 day period from chemo start, identify as regimen 1 (reg1 = ‘1’).
* Svcdates outside the initial 28 will be identified as not regimen 1 (reg1 = ‘0’).
* Also get how many procedures on each svcdate and the largest number of procedures done in regimen 1 timeframe.

Step 3: Get regimen 1 potential start and end dates, next administration date and ordered procedure ranks

* Calculate days into treatment by datediff(svcdate, chemo\_start)
* Identify regimen 1 start and end dates.
* Regimen 1 start date is the earliest date with the most procedures.
* Regimen 1 end date is the last svcdate of regimen 1.
* Also get ranked order of distinct procedure administration and next administration date.

Step 3a: Get drug cocktails

* Group regimen agents by patient and iteration of administration for regimen, and days into treatment. Done in R because ordering a table is not guaranteed in SQL, because calculations may be done on different nodes.

Step 4: Get Max Cycle Length

* Find longest cycle lengths of each regimen’s distinct drugs
* Join drug cocktails onto table so you will still have a claim by claim dataset

Step 5: Find cycle lengths and cycle order

* Cycles lengths are consistently in a multiple of 7 like a week, so cycles will likely either be in a 28 or 21 day cycle.
* If longest cycle is 21 ± 2 then cycle is 21, else 28
* After finding longest cycle of agents in regimen 1, divide the days into treatment for each procedure by the cycle length, to determine what cycle of the regimen it is and round down and add 1. This is to ensure day 0 administrations are counted.
  + Ex: A patient is takes the same drug cocktail at 0, 14, 28, and 42 days into treatment. The maximum cycle length of the agents is 28.

|  |  |  |
| --- | --- | --- |
| Days into Treatment | Cycle Length | Cycle |
| 0 | 28 | 1 |
| 14 | 28 | 1 |
| 28 | 28 | 2 |
| 22 | 28 | 2 |

Step 6: Get final regimen start and dates and the order of each distinct regimen

* Select min and max svcdates partitioned by patient and distinct drug cocktail, which are recognized by regimen start and regimen end
* Get dense\_rank() of each regimen\_start for each distinct patient id to find each distinct regimen
* Find difference between svcdates finalized regimen start dates to find days into each regimen
* Recalculate cycles to have cycles restart for each regimen