

# Incident Rate Analysis

Enhance current incident rate analysis by adding two new measures of incidence. Make available both raw data and visualizations.

**Current state incident rate analysis**, presents a single metric (with stratification) of number of events occurring in a defined population over a specific period of time (numerator), divided by the population at risk for that event over that time (denominator). Two types denominator are supported 1) person at risk, 2) person-time at risk.

	Persons	Cases	Proportion [+ -] per 1k persons	Time At Risk (years)	Rate [+ -] per 1k years
Summary Statistics:	39,173	8,172	208.61	30,032	272.11

Stratify Rule	N	Cases	Proportion [+ -] per 1k persons	Time At Risk (years)	Rate [+ -] per 1k years
1. Is Male	14,943	3,100	207.45	11,464	270.41
2. Is Female	24,230	5,072	209.33	18,569	273.14
3. Age less than 65	1,126	242	214.92	860	281.40
4. Age greater than 65	30,848	6,393	207.24	23,571	271.22

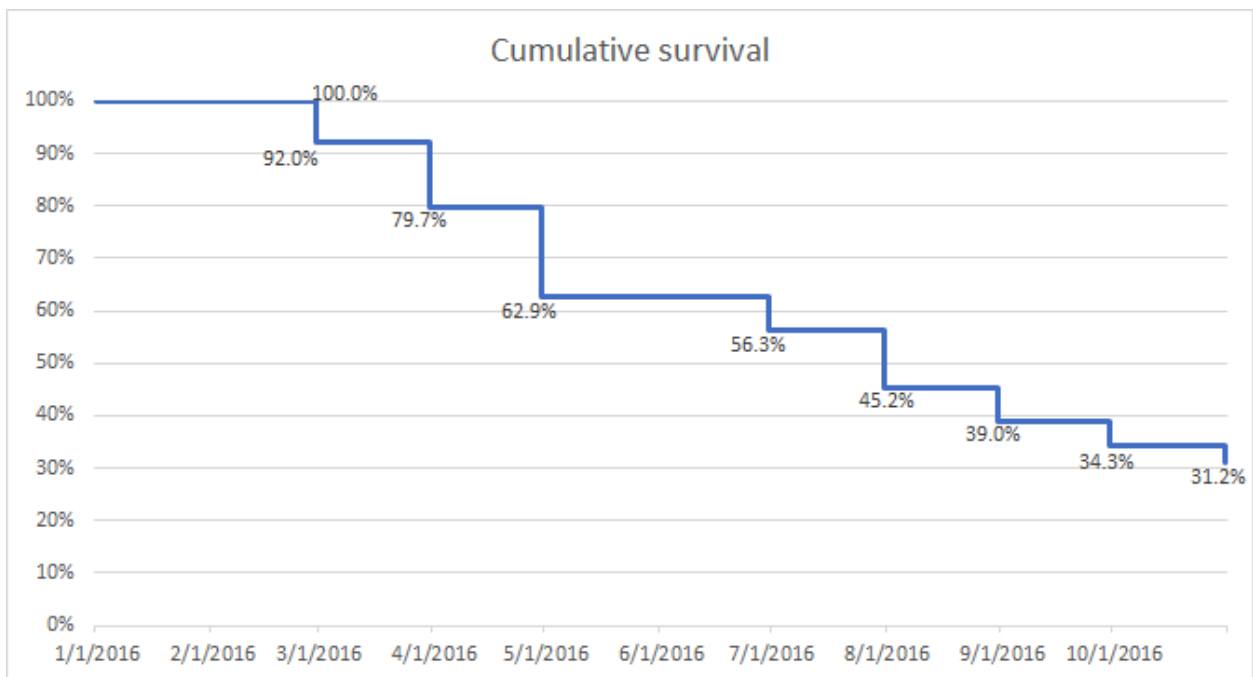
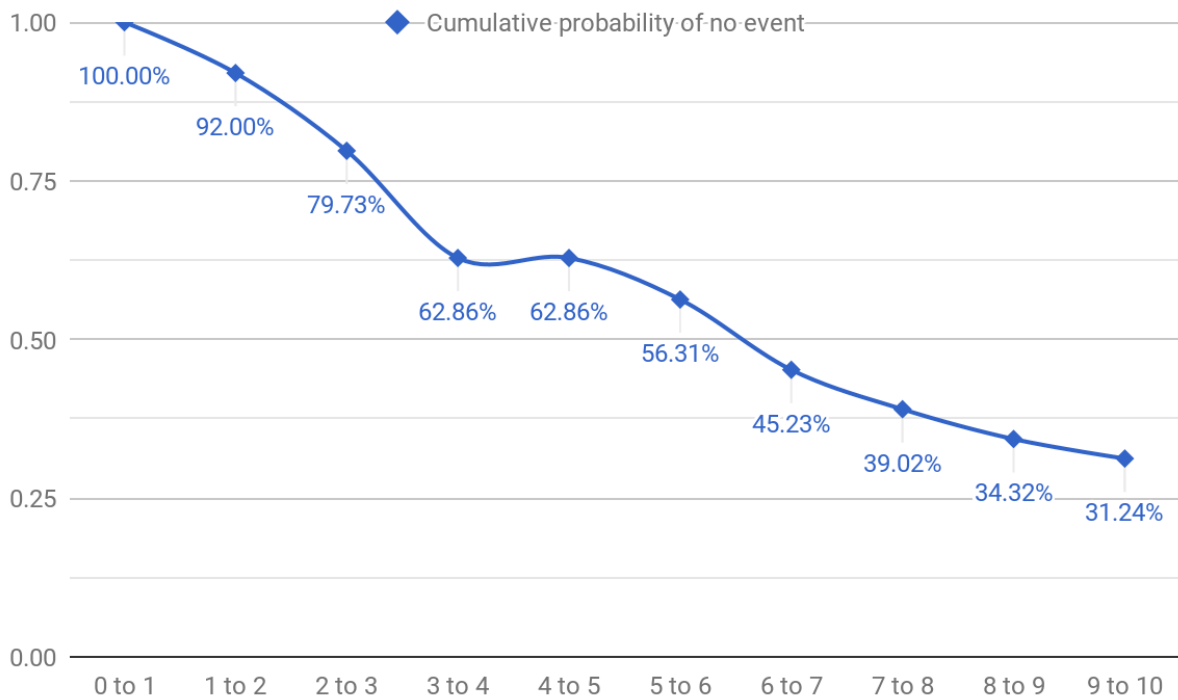
For when denominator = person at risk, the assumption is that follow-up is complete on every individual in the cohort, and does not handle censored observations.

## 1. Add Life table and graph

- Define time periods for life table (e.g., weeks, months, quarters, years)
  - User selects the time-period of interest.
- Application should calculate the following
  - Number at risk at beginning of time period
  - Number of events in each time period
  - Number remaining at risk at end of each time period
  - Cumulative probability of survival at end of each time period
  - Cumulative probability of having event (1-survival) at end of each time period

Exportable as csv

Classic interval life table								
Strata (any strata specified in incident rate analysis)	Time period (w/m/q/y)	At risk in the beginning of period	Censored during the period	Events in period	Probability of event	Probability of no event	Cumulative probability of no event	Cumulative probability of event
None	0 to 1	250	5	20	8.00%	92.00%	100.00%	0.00%
None	1 to 2	225	6	30	13.33%	86.67%	92.00%	8.00%
None	2 to 3	189	1	40	21.16%	78.84%	79.73%	20.27%
None	3 to 4	148	4	0	0.00%	100.00%	62.86%	37.14%
None	4 to 5	144	2	15	10.42%	89.58%	62.86%	37.14%
None	5 to 6	127	0	25	19.69%	80.31%	56.31%	43.69%
None	6 to 7	102	5	14	13.73%	86.27%	45.23%	54.77%
None	7 to 8	83	6	10	12.05%	87.95%	39.02%	60.98%
None	8 to 9	67	4	6	8.96%	91.04%	34.32%	65.68%
None	9 to 10	57	6	10	17.54%	82.46%	31.24%	68.76%



**N at risk:** number of subjects in the target cohort, whose

- target.cohort\_end\_date - target.cohort\_start\_date + 1 >= time-period-start-day, and
- Either
  - outcome.cohort\_start\_date is null or

- $\text{outcome.cohort\_start\_date} > \text{target.cohort\_start\_date}$  AND  
 $\text{outcome.cohort\_start\_date} - \text{target.cohort\_start\_date} + 1 \geq \text{time-period-start-day}$

**N events:** number of subjects in target where

- $\text{target.cohort\_start\_date} < \text{outcome.cohort\_start\_date}$ , AND
- $\text{target.cohort\_end\_date} - \text{target.cohort\_start\_date} + 1 \geq \text{time-period-start-day}$ , AND
- $\text{outcome.cohort\_start\_date} - \text{target.cohort\_start\_date} + 1$  between  $\text{time-period-start-day}$  and  $\text{time-period-end-day}$

**N censored:** number of subjects in target where

- $\text{Outcome.cohort\_start\_date}$  is null, AND
- $\text{target.cohort\_end\_date} - \text{target.cohort\_start\_date} + 1 < \text{time-period-start-day}$

I think this should be:

$\text{target.cohort\_end\_date} - \text{target.cohort\_start\_date} + 1 \geq \text{time-period-start-day}$  AND

$\text{target.cohort\_end\_date} - \text{target.cohort\_start\_date} + 1 \leq \text{time-period-end-day}$

**Pseudo code:**

1. Check if target and outcome cohorts have been generated
2. Use first outcome per  $\text{subject\_id}$  only where  $\text{target.cohort\_start\_date} < \text{outcome.cohort\_start\_date}$
3. Generate period reference table
  - a. Depends on user selection of time-period (w, m, q, y) - parameterized
  - b. Calculate range between  $\max(\text{cohort\_end\_date}) - \max(\text{cohort\_start\_date})$  and use it to calculate needed range for period reference table
4. Use period reference table for group by, calculate
  - a. N at risk
  - b. N with event
  - c. N censored
  - d. Compute probability and cumulative probability

## 2. K-M curves

- K-M is similar to Life-table, but instead of user-defined equal intervals of time, probability of the event is calculated as the event happens.
- Time-periods are always days
- Should be informed by more granular information than what is displayed in the life table (i.e., it should use information by **day** not by weeks, months, etc.)
- Create this granular table but don't display to user unless they want to see it
- K-M curve then is simply time (in days) on the x-axis and cumulative probability of survival on the y-axis.
- UI should have cursor hover feature to show cumulative survival at a point on curve
- Could also have option to show cumulative probability of **event** rather than survival. This is just 1-survival and the graph will start at 0 and go up instead of starting at 1.0 and going down.