## 1 Model-based Risk (and Cumulative Incidence) estimates; Number Needed to Treat.

i. In "Rosuvastatin to Prevent Vascular Events in Men and Women with Elevated C-Reactive Protein" (2008 NEJM), Fig 1 contained a plot of the 2 K -M-based cumulative incidence (risk) curves that ended (as follow-up did) at 4.5 years. The difference in the 1,2 and 4 -year risk estimates were used to calculate NNTs. For example, based on the 4 -year risks read off the K-M curves, the number needed to treat for 4 years is 31 .

The authors also 'went beyond their data' and calculate 5-year risks:

If 4 -year risks are projected over an average 5 -year treatment period, as has been commonly done in previous statin trials according to the method of Altman and Andersen ${ }^{22}$, the NNT to prevent the occurrence of one primary end point is 25 .

Last year, we calculated that in the control arm, the incidence density of the primary endpoint was (approx.) 1.25 events per 100 PY in the first 2 years, and 2.0 events per 100 PY in years 3 and 4 . Use these incidence densities, and assuming an incidence density of 2.5 events per 100 PY in year 5 , to project the 5 -year risk for similarly untreated subjects. Hint: the answer is NOT (or at least not exactly) 9\%. Use the rel'n $b / w$ the integral of the incidence density fn. and the survival fraction.

$$
\begin{aligned}
5 \text {-year Risk } & =1-\exp \left[-\int_{0}^{5} I D(t) d t\right] \\
& =1-\text { PoissonProb }[0 \text { events } \mid \text { Expected No. events }=\mu]
\end{aligned}
$$

$\mu=0.0125 \times 2+0.02 \times 2+0.025 \times 1$
$=0.09$ events, if 1 [not necessarily same] person at risk for 5 years [the definition of 5 person years is 'always 1 person on duty' for 5 years]

$$
\begin{aligned}
5 \text {-year Risk } & =1-\exp [-0.09] \\
& =1-0.914 \\
& =1-\text { 'survival' probability } \\
& =0.086 \\
& =8.6 \%
\end{aligned}
$$

ii. In "Extended work duration and the risk of self-reported percutaneous injuries in interns' by Ayas et al. in JAMA. 2006;296:1055-1062, Table 1 reports that the percutaneous Injury (PI) rate for obstetrics/gynecology (OB/GYN) residents was 0.0975 injuries/Intern-Month (I-M).

Using this incidence rate, calculate the probability that an average-risk ob/gyn resident would have no (or the complement, at least one), percutaneous Injury by the end of (a) 1 month (b) 12 months of experience? i.e. what is the probability of 'surviving' these amounts of experience without a PI? The complement is often referred to as 'cumulative incidence' or 'risk'. Hint: draw the (in this case constant over time) incidence density function, obtain its integral over the time-span in question, and use the formula linking the integral and cumulative incidence.
(a)

$$
\begin{aligned}
1 \text {-month Risk } & =1-\exp \left[-\int_{0}^{1} I D(t) d t\right] \\
& =1-\text { PoissonProb }[0 \text { events } \mid \text { Expected No. events }=\mu]
\end{aligned}
$$

$I M=$ Intern-Months
$\mu=0.0975 \times 1 \ldots \ldots .$. [notice the UNITS: $\frac{\text { No.events }}{I M} \times M=$ No. events]
$=0.09$ events, if 1 [not necessarily same] person at risk for 1 mo. [the definition of 1 intern month is 'always 1 intern on duty' for 1 mo .]

SO, ...

$$
\begin{aligned}
1 \text {-month Risk } & =1-\exp [-0.0975] \\
& =1-0.9071 \\
& =0.0929 \\
& =9.29 \%
\end{aligned}
$$

(b)

$$
\begin{aligned}
\text { 12-month Risk } & =1-\exp [-0.0975 \times 12] \\
& =1-\exp [-1.17] \\
& =1-0.31 \\
& =0.69 \\
& =69 \%
\end{aligned}
$$

iii. In (i) the cumulative incidence was close to the integral of the incidence function, while in (ii) it was not. Why? Develop/otherwise find a rule-ofthumb for when the approximation: 'CI $\approx$ integral of incidence function' is accurate to 1 (and 2) significant figures?

| integral cum.inc |  |
| :---: | :---: |
|  |  |
| 0.001 | 0.001 |
| 0.010 | 0.010 |
| 0.020 | 0.020 |
| 0.030 | 0.030 |
| 0.040 | 0.039 |
| 0.050 | 0.049 |
| 0.060 | 0.058 |
| 0.070 | 0.068 |
| 0.080 | 0.077 |
| 0.090 | 0.086 |
| 0.100 | 0.095 |
| 0.110 | 0.104 |
| 0.120 | 0.113 |
| 0.130 | 0.122 |
| 0.140 | 0.131 |
| 0.150 | 0.139 |
| 0.160 | 0.148 |
| 0.170 | 0.156 |
| 0.180 | 0.165 |
| 0.190 | 0.173 |
| 0.200 | 0.181 |

By inspection of the two columns, we can see that the first 2 significant figures agree up to an integral of 0.03 , and first 1 significant figure is the same up to an integral of 0.19.
This is because the mathematical function $\exp [-x]$ can be approximated by $1-x$ when $x$ is small, and so $1-\exp [-x] \approx x$ when $x$ is small.

## 3 Summary measures

The abstract of the COMPARE study states "The primary endpoint occurred in $56(6 \%)$ of 897 patients in the everolimus-eluting stent group versus $82(9 \%)$ of 903 in the paclitaxel-eluting stent group (relative risk 069 [ $95 \%$ CI $0.50-$ $0.95]$, p value for superiority=0.02)."
i. Ignore the small amount of censoring, treat the 56 of 897 and 82 of 903 as binomial proportions (risks), and use Rothman's formula for the interval estimate for a relative risk to come up with your own $95 \%$ interval estimate for the ( 1 year) relative risk
The formula is $7-3$ on p 135 of Rothman2002. The key is to work in the $\log$ scale: the log of a relative risk is the difference of the logs of the risk estimates in the index and reference risks, and the variance of this diff. is the sum of the variances of the logs of these. i.e., with $p=\frac{\text { no. cases }}{n}$,
$\operatorname{Var}[\log [\widehat{r i s k}]]=\operatorname{Var}[\widehat{r i s k}] \times\left[\frac{\log [\text { risk }]}{d \text { risk }}\right]^{2}=\frac{p q}{n} \times\left[\frac{1}{p}\right]^{2}=\frac{1}{\text { no. cases }}-\frac{1}{n}$
So, with subscripts of 1 and 0 for the index and reference categories,

$$
\begin{gathered}
\operatorname{Var}[\log [\text { riskRatio }]]=\left\{\frac{1}{\text { no. cases }} 1\right. \\
\left.-\frac{1}{n_{1}}\right\}+\left\{\frac{1}{\text { no. cases }} 0\right. \\
\left.-\frac{1}{n_{0}}\right\} \\
\text { Here }: \operatorname{Var}[\log [\text { riskRatio }]]=\left\{\frac{1}{56}-\frac{1}{897}\right\}+\left\{\frac{1}{82}-\frac{1}{903}\right\}=0.1668^{2} \\
\qquad r=\frac{56 / 897}{82 / 903}=0.69
\end{gathered}
$$

$95 \%$ multiplicative M.E for riskRatio : $\exp [1.96 \times 0.1668]=1.39$
$95 \%$ CI for riskRatio : $0.69 \div 1.39$ to $0.69 \times 1.39$, i.e., 0.50 to 0.95 .
ii. Suppose that an economic analysis of the new stent used a 5 -year horizon rather than a 1 year one. For days 1-10, use the incidence densities seen in the first 10 days; for the reminder of the first year, use those seen in days 11-360; For years 2 to 5 use those seen in days 11-360. From these inputs, calculate 5 year risks, a 5 year risk difference, and the NNT to avoid 1 (primary endpoint) event over these 5 years. Comment on the appropriateness of the single summary 'relative risk at 1 year' measure the authors reported. You might want to read Miguel Hernan's article on the hazards of proportional hazards (i.e. proportional incidence densities!).

- The first computing step in computing the 4 rates (for days 1-10 and 11360 in each arm) is to use (say) the the survSplit function in the survival package to split each patient's time into 2 portions.
> library(survival)
Loading required package: splines
$>$ ds\$day $=$ ds\$day+ 1
> \# to track the time-slices for each person, give each patient an ID ds $\$$ ID $=1:$ length(ds $\$$ day)
$>$ cutpoints $=c(0.5,10.5)$ \#\# for 2 intervals $0-10$ and $10-360$
$>$
> split=survSplit(ds,cut=cutpoints,end="day",event="event",start="start.day")
> split\$pt.days = split\$day- split\$start.day
> head(split)
tx day event ID start.day pt.days

| 1 | 0 | 0.5 | 0 | 1 | 0 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| 2 | 0 | 0.5 | 0 | 2 | 0 |
| 3 | 0 | 0.5 | 0 | 3 | 0 |
| 4 | 0 | 0.5 | 0 | 4 | 0 |
| 5 | 0 | 0.5 | 0 | 5 | 0.5 |
| 6 | 0 | 0.5 | 0 | 6 | 0 |

$>$ tail(split)


| 5395 | 1 | 361 | 0 | 1795 | 10.5 |
| :--- | :--- | :--- | :--- | :--- | :--- |
| 350.5 |  |  |  |  |  |


| 5396 | 1 | 361 | 0 | 1796 | 10.5 |
| :--- | :--- | :--- | :--- | :--- | :--- |

$\begin{array}{llllll}5397 & 1 & 361 & 0 & 1797 & 10.5 \\ 350.5\end{array}$
$\begin{array}{llllll}5398 & 1 & 361 & 0 & 1798 & 10.5 \\ 350.5\end{array}$
$\begin{array}{llllll}5399 & 1 & 361 & 0 & 1799 & 10.5 \\ 350.5\end{array}$
$10.5 \quad 350.5$

- Now aggregate (sum) the person-time \& events in each of these windows $\times 2$ arms
events.interval.tx=aggregate(split\$event,
by=list(tx=split\$tx,mid.interval=split\$start.day+5), FUN="sum")
$>$ PT.interval.tx=aggregate(split\$pt.days,
by=list(tx=split\$tx,mid.interval=split\$start.day+5),FUN="sum")
> events.interval.tx
tx mid.interval $x$
1050
$2 \quad 5.00$
$30 \quad 5.533$
$\begin{array}{lll}4 & 1 & 5.518\end{array}$

| 5 | 0 | 15.549 |
| :--- | :--- | :--- | :--- |
| 6 | 1 | 15.538 |

> PT.interval.tx
tx mid.interval $x$

| 1 | 0 | 5.0 | 451.5 |
| :--- | :--- | :--- | :--- |


| 2 | 1 | 5.0 | 448.5 |
| :--- | :--- | :--- | :--- |


| 3 | 0 | 5.5 | 8789.5 |
| :--- | :--- | :--- | :--- |

$1 \quad 5.5 \quad 8811.0$
$1 \quad 15.5295507 .0$
$>$ dta= cbind(PT.interval.tx,events.interval.tx\$x)
> names(dta) [3]="Pt.days"
$>$ names(dta) [4] ="events"
$>\mathrm{dta}$
tx mid.interval Pt.days events

| 1 | 0 | 5.0 | 451.5 | 0 |
| ---: | ---: | ---: | ---: | ---: |
| 2 | 1 | 5.0 | 448.5 | 0 |
| 3 | 0 | 5.5 | 8789.5 | 33 |
| 4 | 1 | 5.5 | 8811.0 | 18 |
| 5 | 0 | 15.5 | 295507.0 | 49 |
| 6 | 1 | 15.5 | 300830.5 | 38 |

- Now ready to compute the 4 rates...
> rates = dta[,"events"] / dta[,"Pt.days"];
> rates
[1] 0.00000000000 .00000000000 .00375447980 .00204290090 .0001658167 [6] 0.0001263170
rate. $0=\mathrm{c}($ rates [3], rates [5]) ; rate. 0
[1] 0.00375447980 .0001658167
rate. $1=\mathrm{c}$ (rates[4], rates[6]) ; rate. 1 <<<<<<<<<<<<<<<<<<<<<<<<<
[1] 0.00204290090 .0001263170
- Compute the 2 5-year integrals, and from them the 25 -year risks...
extra.years=4;
> interval.widths=c(10, 355, extra.years*365) <<<<<<<<<<<<<<<<<<<<<
\# 1st 10d, rest of $y 1, y 2-5$
$>$
> integral. $0=$ sum(c(rate.0,rate.0[2]) * interval.widths) ; integral.0
[1] 0.3385021
> integral.1=sum(c(rate.1,rate.1[2]) * interval.widths) ; integral.1
[1] 0.2496943
$>$
> FiveYearRisk. $0=1-\exp (-i n t e g r a l .0)$; FiveYearRisk. 0
[1] 0.2871627
> FiveYearRisk. 1 = 1 - exp(-integral.1) ; FiveYearRisk. 1
[1] 0.2209611
FiveYearRiskDiff = FiveYearRisk. 0 - FiveYearRisk.1; FiveYearRiskDiff
[1] 0.06620162
> NNT = 1/FiveYearRiskDiff ; round(NNT,0)
[1] 15

