Missing Data/ Time Dependent Covariates

21 October 2014
Dave Glidden
Biostatistics 210

Maximum Likelihood (ML)

Based on implication #1

- Maximum likelihood valid for MAR data
  as long as CMAR variables incorporated
- Works especially well for missing outcome
  especially in longitudinal analysis
  where attrition depends on previous values
- ML often fit with the EM algorithm
  EM also iterative fills in missing data
- Imputation & analysis models consistent
  MI and maximum like very similar

Disadvantages of ML

- Cumbersome for missing predictors
  don't usual model their distribution
- Awkward if CMAR variables are not a part of desired
- Lacks flexibility of MI

Inverse Weighting

Based on implication #1

- Model pr(non missingness|C_{MAR})
  weight by chance of being observed
- Two step approach
  not implemented in software
  easy to mess up (my experience)
- Clear connection to complete data analysis
- Inefficient

3 Methods

- Same MAR assumption
- Inverse weight:
  model missingness, data model flexible
- MI/ML:
  model data, missingness model unspecified

Recommendations

- Investigate reasons for missing data
  gain insights into modeling and assumptions
- Approach using MAR assumption
  MCAR is simply too restrictive
- MI is preferred approach to missingness
  chained equation gives most flexibility
- Missing only data in outcome use ML
  unless you'd need to adjust for mediators
  then use MI

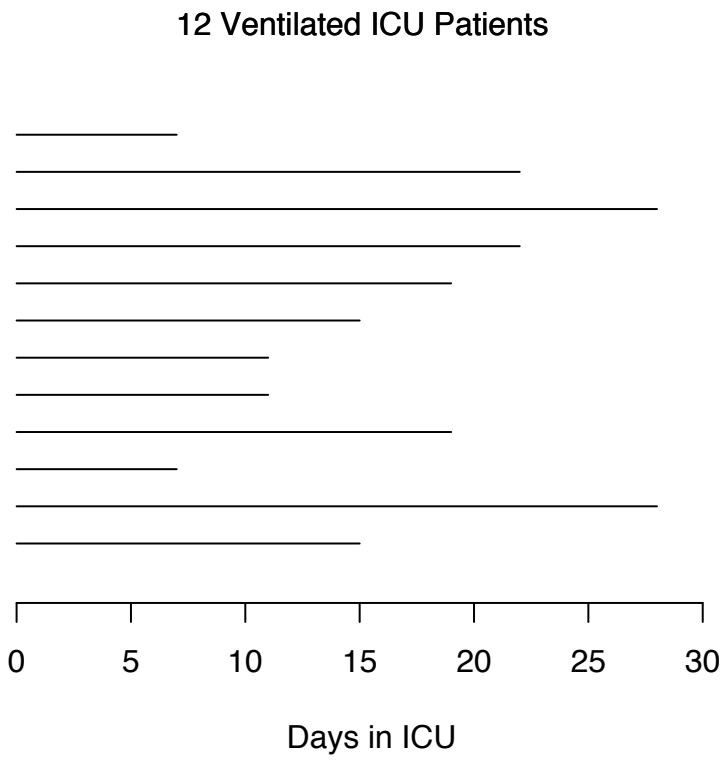
Recommendations

- Missing data in predictors, use MI
- Missing data in only in outcome
  - if MAR vars in stat. model, complete case *as long as ML based method*
  - MAR vars not in stat. model, use MI
  - If want to use GEE, use MI

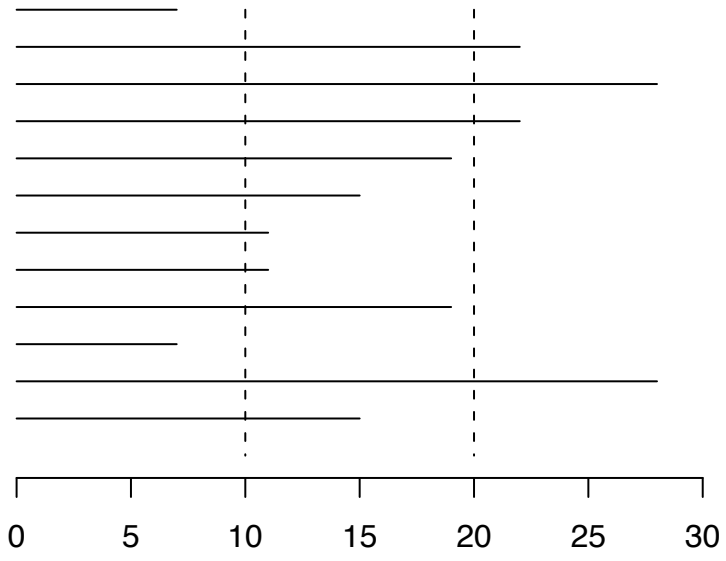
Time Dependent Covariates

Data Example

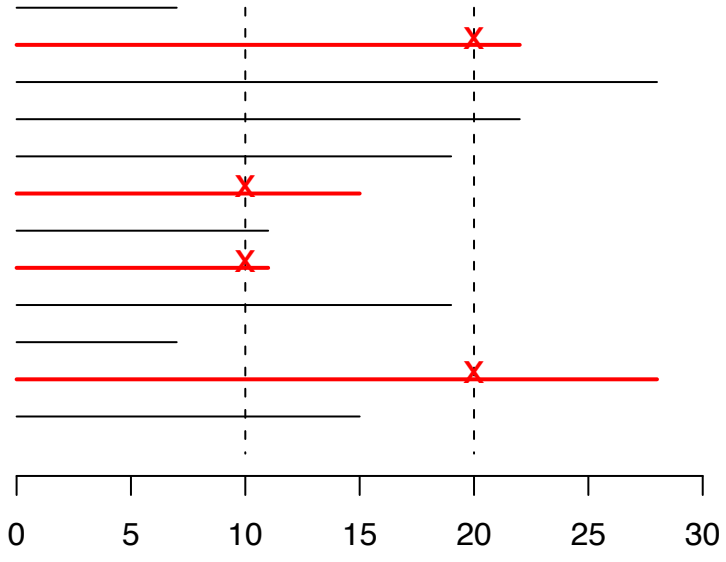
- Collaboration with ICU investigators
- Cohort of ventilated patients
- Ventilator assisted pneumonia (VAP)
- Does it extend time on the ventilator? Does it increase risk of death?
- VAP develops over clinical course in ICU

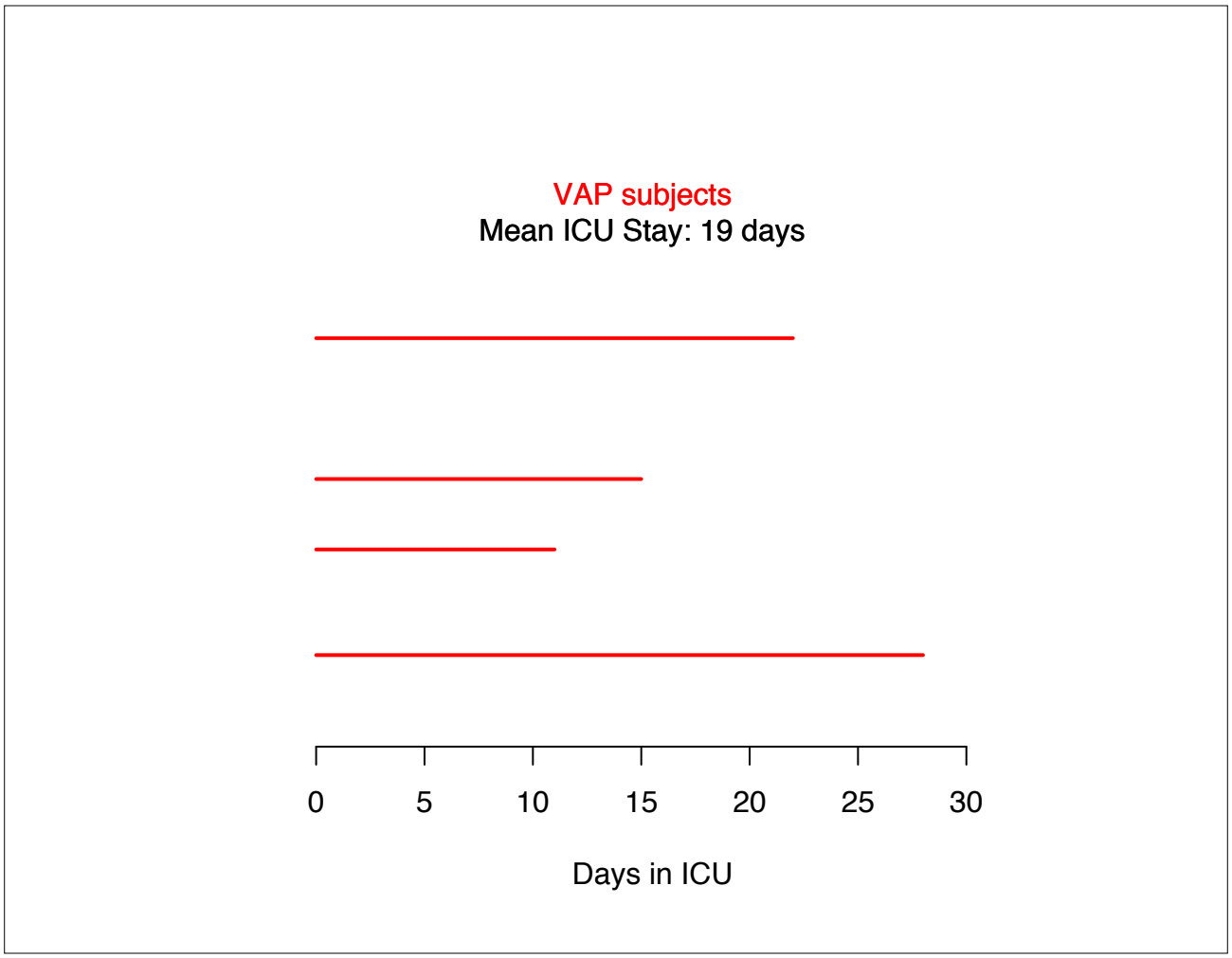


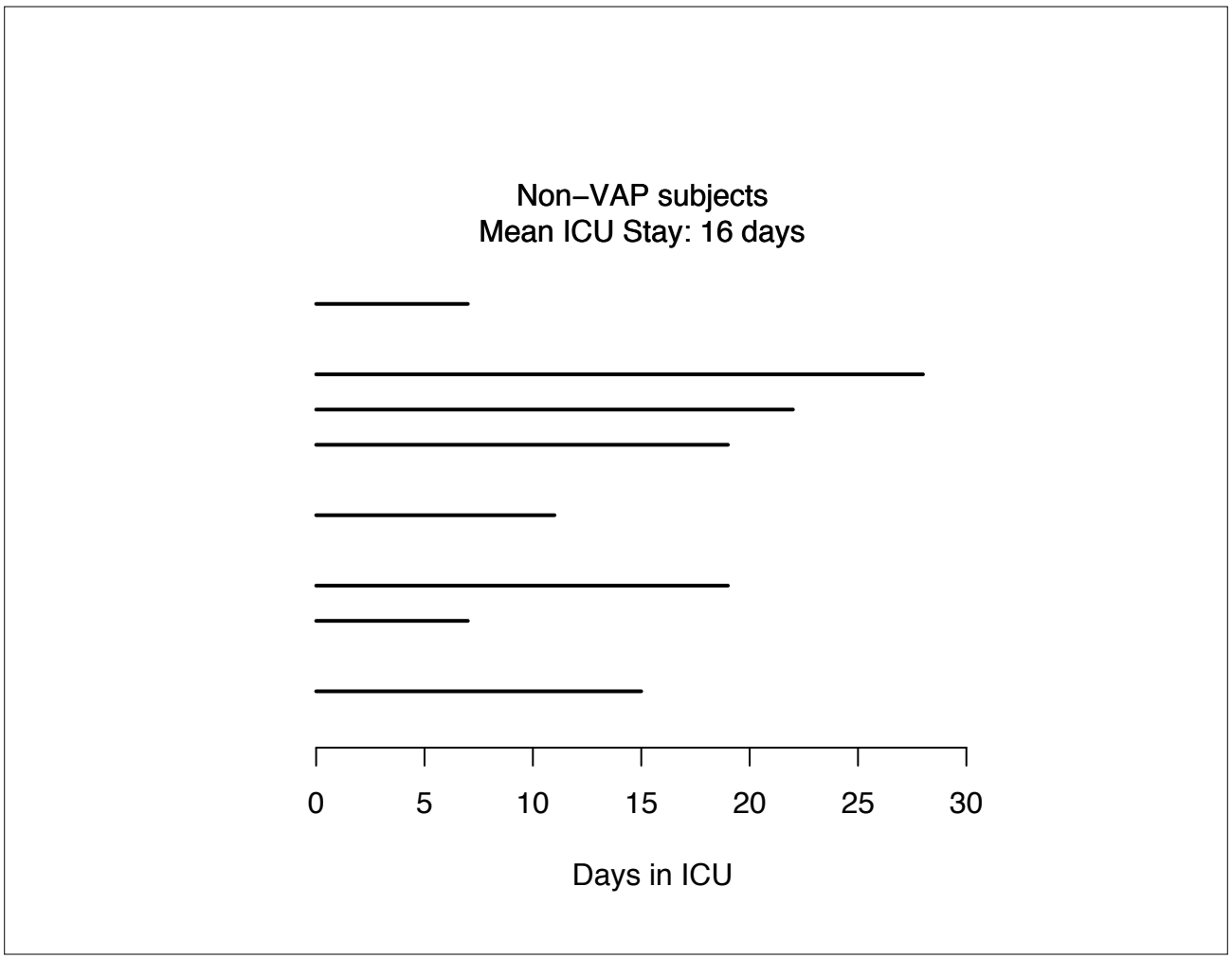
VAP Occurs at 10 or 20 days

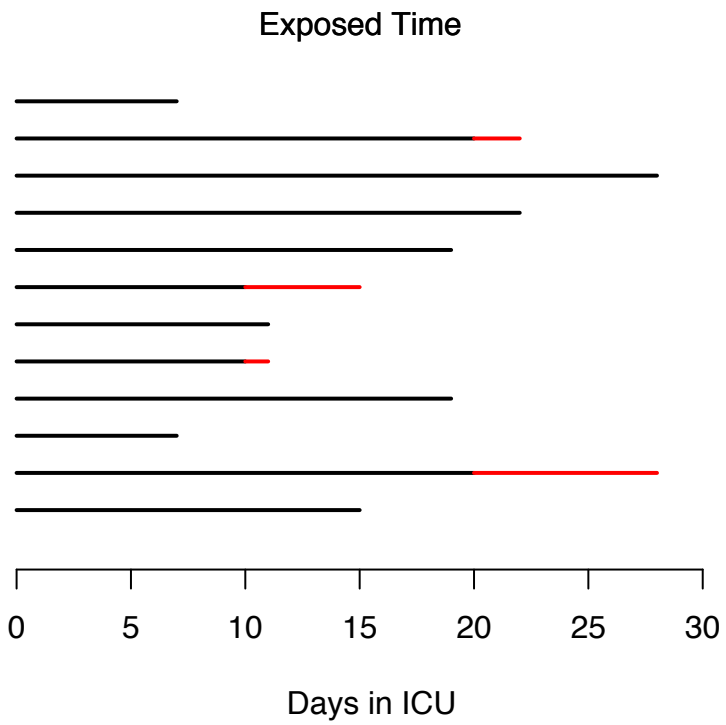


VAP Occurs at 10 or 20 days
Subjects Randomly Selected for VAP




VAP subjects
Mean ICU Stay: 19 days
0
5
10
15
20
25
30
Days in ICU


Non–VAP subjects
Mean ICU Stay: 16 days
0
5
10
15
20
25
30
Days in ICU



The Bias

- Patients ever exposed to VAP vented longer
- VAP patients: two kinds of vent time
- Time before VAP: longer than non-VAP
  the so-called immortal time
- Time after VAP: same as VAP- subjects
- Should time before VAP count toward VAP+?

Other Examples

- PTLD increase risk of death in kidney tx?
- Does heart tx extend life in listed patients?
- Do OSCAR winners live longer?
- Common thread: exposure occurs sometime after follow-up
- PTLD, transplants, awards occur over time

SCC and voriconazole

- Solid organ transplant recipients
- Immune suppression
- Given voriconazole for fungal prophylaxis
- voriconazole introduced ~10 years ago
- Recent reports: may be associated with onset of squamous cell carcinoma (SCC)

UCSF Retrospective Cohort

- 380 Patients with Lung Transplant at UCSF
- Between 23 Oct '91 and 26 Dec '10
- Followed through mid 2011
- 55 patients developed SCC during FU
  17% in those with voriconazole
  9% in those with no voriconazole

Patient 1123

Example

- Transplanted 4/25/05
- SCC free and alive on 10/31/06
  554 days after
- 4/28/05 - 6/16/05: 49d @ 400mg/day =
  49*400/(400*365) = 0.135
- 9/7/05 -11/10/06: @ 400mg/day
  419 days thru 10/31 = 1.148
- Cumulative exposure = 1.282 year of 400mg
- Need this for every person

Analysis Framework

- Cox model for rate of SCC
- Predictor: exposure to voriconazole
- Bad way:
  - fixed predictor
- Better way
  - break up follow up: td predictor

Fixed Predictor

abpatientcode	time	event	evervori	totaldose	ageattp
1123	837	2	1	1,36	64

Ever Voriconazole

fixed time covariate

```
No. of subjects =          379                     Number of obs   =      8433
No. of failures =           54
Time at risk    =       413146
                                                   LR chi2(1)      =      3.47
Log likelihood  =   -250.63795                     Prob > chi2     =    0.0625

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
evervori_f~d |   1.787881   .5859601     1.77   0.076     .9405156    3.398689
------------------------------------------------------------------------------
```

ever voriconazole versus never

Dose of Voriconazole

fixed time covariate

```
Cox regression -- Breslow method for ties

No. of subjects =          379                     Number of obs   =      8433
No. of failures =           54
Time at risk    =       413146
                                                   LR chi2(1)      =      0.28
Log likelihood  =   -252.23301                     Prob > chi2     =    0.5970


------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   totaldose |   1.051435   .0973257     0.54   0.588     .8769829    1.260589
------------------------------------------------------------------------------
```

per year at 400 mg dose

Time-Dependent Covariate

TDC

- Predictors whose value may vary with time
- And measured as a part of the study
- Longitudinal data in a predictor
- Outcome is usually longitudinal
  GEE or a mixed model
- Or as a part of a Cox model

Td predictor

abpatientcode	from	to	event	cumdose	ageattp
1123	0	91	0	0.13	64
1123	91	198	0	0.31	64
1123	198	263	0	0.48	64
1123	263	398	0	0.85	64
1123	398	405	0	0.87	64
1123	405	447	0	0.99	64
1123	447	545	0	1.26	64
1123	545	554	0	1.28	64
1123	554	590	0	1.31	64
1123	590	637	0	1.36	64
1123	637	638	0	1.36	64
1123	638	723	0	1.36	64
1123	723	761	0	1.36	64
1123	761	800	0	1.36	64
1123	800	817	0	1.36	64
1123	817	837	2	1.36	64

The Cox Model

with time-dependent covs

- $h(t|x_1,..,x_p) = h_0(t)\exp(\beta_1 x_1 + ... + \beta_p x_p)$
- Model hazard (like a rate) function
- baseline hazard not specified
- Predict effect acts multiplicatively on hazard
- $h(t) = h_0(t)\exp(\beta_1 x_1\ (t) + .. + \beta_p x_p\ (t))$
  allows the predictors to vary with time

Why use time-dependent covariate?

- Factor or exposure is time varying
- Can increase predictive power
  e.g., viral load in HIV
- Aligning timing of exposures & outcomes
  avoiding immortal time bias

Defining Exposure

- Created exposure variable:
  x(t) dose-days of voriconazole through t
- Acknowledges that exposure changes
- Time prior to exposure, x(t)=0
- After exposure accumulates through t
- If exposure stops: x(t) doesn't increase

Time-Dependent Covariate

- Data management is biggest hurdle
- Need a long record dataset for each person
- Need know for series of time periods
  - Did person get SCC or get censored?
  - What was value of td cov for that period?
- Series of time points -- observed times of SCC

Td predictor

abpatientcode	from	to	event	cumdose	ageattp
1123	0	91	0	0.13	64
1123	91	198	0	0.31	64
1123	198	263	0	0.48	64
1123	263	398	0	0.85	64
1123	398	405	0	0.87	64
1123	405	447	0	0.99	64
1123	447	545	0	1.26	64
1123	545	554	0	1.28	64
1123	554	590	0	1.31	64
1123	590	637	0	1.36	64
1123	637	638	0	1.36	64
1123	638	723	0	1.36	64
1123	723	761	0	1.36	64
1123	761	800	0	1.36	64
1123	800	817	0	1.36	64
1123	817	837	2	1.36	64

Cox Model w/ TD Cov

- Patient #37: SCC 554d after transplant
  received 1.01 years of voriconazole
- Think of him as a "case"
  to be compared to a set of controls
  matched on follow-up
- More vori exposure than typical case?
- How much vori in those SCC-free 554 days?

stset

- stset to, failure(event==1) id(abpatientcode)
- "from" is not need but helps me keep track
- event: 0 censored, 1 SCC, 2 death
  if want Cox on SCC, set event to 1
- Must use id for multiple events data
- Then just apply Cox
- Not good for anything else!
  multiple records foul all else up

Ever Voriconazole

as a time-dependent covariate

```
stcox evervori_td

Cox regression -- Breslow method for ties

No. of subjects =          379                     Number of obs    =      8433
No. of failures =           54
Time at risk    =       413146
                                                    LR chi2(1)       =      8.53
Log likelihood  =    -248.1054                      Prob > chi2      =    0.0035

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 evervori_td |   2.444013   .8028329     2.72   0.007     1.283786      4.6528
------------------------------------------------------------------------------
```

Dose of Voriconazole

as a time-dependent covariate

```
stcox cumdose

No. of subjects =          379                     Number of obs    =      8433
No. of failures =           54
Time at risk    =       413146
                                                    LR chi2(1)       =      5.93
Log likelihood  =    -249.4083                      Prob > chi2      =    0.0149

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     cumdose |   1.311188   .1346107     2.64   0.008     1.072205    1.603438
------------------------------------------------------------------------------
```

per year at 400 mg dose

TD Covs Eliminates ITB

- Behind the scenes it's implementing the matching approach
  comparing comparable follow-up
  looking at effects after exposure
- But using all the data in a flexible way
  no artificial matches
  all available data used

Sparsity

- HIV chemoprophylaxis
  with anti-viral drug emtricitabine+tenofovir
- Risk of HIV high associated with levels
  pill taking fluctuates, drugs levels per 24w
- Week 24: 10.5 fmol/10^6 cells
- Week 48: 10.5 fmol/10^6 cells
- What is the level at Weeks 32, 54?

Paradox

reverse causation in smoking exposure

- Coronary artery surgery study
  smoking status per 6 months
- Current smoking: reduced risk of death
  not statistically significant
- Most deaths among smokers
  but had stopped smoking just prior to death
- Hospitalization/CHF lead quitting before death
- Their soln: lag smoking by 6 month

Histories

- BCC study: voriconazole hx over many years
- Which part is relevant for risk of BCC?
  current voriconazole
  current dose
  total accumulated voriconazole
  accumulated exposure in last year
- Combine subject-matter knowledge
  and can see what fits better

Ever vs. Cumulative Voriconazole

```
. stcox cumd evervori_td

No. of subjects =          379                     Number of obs   =      8433
No. of failures =           54
Time at risk    =       413146
                                                    LR chi2(2)      =      9.72
Log likelihood  =   -247.51186                      Prob > chi2     =    0.0077

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     cumdose |   1.155081   .1469683     1.13   0.257     .9001372    1.482232
 evervori_td |   2.026276   .7556136     1.89   0.058     .9756167     4.20841
------------------------------------------------------------------------------
```

ever vs. cumulative -- ever appears stronger

Three Things to Think Of

1. May be gaps in exposure (sparsity)
2. Can introduce additional confounding
   some td covs can even be mediators
3. Have to distill a complex history
   rich modeling choices

Baseline Covariates

- Osteoporosis study
- Baseline BMD predicts fracture over 3 years
- That, in itself, is powerful
- Prognostic over many years powerful
- Baseline and td cov analyses complimentary
- A baseline value can be just fine

Recognizing ITB

- Carefully examine when and how people enter a "group" or accrue their exposure
- Is their a time-related requirement?
  if yes, ITB
- Based on classification @ time 0 or later?
  if later, ITB
- TD covs can account for this

TD Covs

- Clarifying exposure and events
- Addresses immortal time bias
- Dataset construction can be tedious
- Fitting model is easy
- Mindful of inverse associations and mediators
- Rich choices for incorporating hx