Overview.

Summarize events and follow-up time
- stptime

Figures -- Can describe as “failure” w/event (vs “survival” w/o event)
- sts graph

Summarize specific timepoints
- sts list – Obtain failure (or surv) probabilities, get 95% CIs for the probabilities
- stci – Obtain mean, median, percentile failure (or surv) times

Test of KM curves (unadjusted)
- Log-rank test – sts test

Regression (unadjusted and adjusted)
- Continuous time
  - Cox regression (Proportion hazards) -- stcox
- Discrete time
  - Proportional odds -- logit
  - Proportional hazards -- cloglog

Fixed at baseline versus Time-varying covariates

Assessing the proportional hazards assumption (time-dependent effects)
- stphtest, stphplot, stcoxkm

Continuous time -- Parametric fitting of baseline hazard (ex Exponential, Weibull)
- Relationship to Poisson, Negative Binomial

Multiple/recurrent events
- Mixed effects (frailty), GEE
- Anderson-Gill, Poisson, Negative Binomial
- (Paired: IMCI vs PCR; CTL; Analyst: Lara Diener)

1. Sample size/power calculations

Constant hazard (exponential)
Proportional hazards (HR constant over time)

sampsi, stpower in Stata

2. Data Management/Preparation

Construction of time to event variables

Treating time as *continuous:*

Wide format: One record per subject

```
stset studyday, failure(cd4_lt350)
stptime, by(arm) per(100)
```

Long format: One record per visit per subject
*Can also examine time-varying covariates

```
stset studyday, failure(cd4_lt350) id(studyid)
```

Kaplan-Meier plot of time to disease progression by cohort, with risk table

```
sts graph, by(arm) riskt(,failevents order(2 "Control" 1 "Intervention") ///
       title("Number at risk (Events) " , size(3) ) size(3)  rowtitle(, justification(left)) ) ///
       scheme(s2mono) plotregion(style(none) fcolor(white)) graphregion(fcolor(white)) ///
       title("") ytitle("Probability") xtitle("Time From Enrollment, Years") ///
       legend(ring(0) rows(2) pos(5) order(2 1) label(1 "Intervention") ///
       label(2 "Control") region(lstyle(none) color(white) ) )
```

List of the K-M survivor (or failure) function
Obtain the estimated survival probabilities, and 95% CIs

```
sts list, by(arm)
```

Obtain the estimated times at which failure probability is 10%, 50%, etc.

```
stci, p(10) by(arm)
stci, p(50) by(arm)
```

Fit the Cox regression model to estimate the Hazard Ratio, CI, and p-value

```
stcox arm
```

Checking the proportional hazards assumption

1. **Schoenfeld residuals: goodness-of-fit tests and plots**

- The global test tests the null hypothesis that there are no proportional hazards violations among the variables in the model. (p value < 0.05 suggests one or more violations)
  - Individual tests test null hypotheses for each variable
  - Doing one global test is probably adequate for model-checking
- If the global test suggests proportional hazards violations, consider further diagnostics.
- Can also plot Schoenfeld residuals vs. time to look for increasing or decreasing trends that are suggestive of proportional hazards violations.

```
stcox arm cd4enrl
estat phtest, log detail
estat phtest, log plot(armn) yline(0)
```

2. **Log-log plots**

- Under a proportional-hazards assumption, the separation between –log-log plots should be constant over time.
- Do the transformed survival curves (i.e. –log-log plots) for each of the groups look approximately parallel? (This is a judgment call)
  - Crossing of the curves can indicate trouble.

```
stphplot, strata(armn) adj(cd4enrl)
```

3. **Observed survival (from Kaplan-Meier estimates) vs. fitted survival (from Cox model)**

- Compare Kaplan-Meier estimates to fitted survival curves from Cox regression
- If model is a good fit, then K-M estimates should be close to fitted estimates

```
stcoxkm, by(armn)
```

4. Use a stratified Cox model if possible.

- The stratified Cox model allows the form of the baseline hazard function to vary across levels of stratification variables.
- Allows a factor to be adjusted for without estimating its effect on the outcome.
- Use for predictors that don't satisfy proportional hazards assumption, where you don't need to estimate the predictor's association with outcome. (i.e. an adjustment variable, rather than an exposure of interest)
  - e.g. Variable "toilet type" in the current example

```
stcox armn cd4enrl, strata(stratavar)
```

5. Include a time-dependent effect (to estimate a time-varying hazard ratio):

- This is a covariate-by-time interaction.
  - The coefficient estimates the change in the hazard ratio as a function of time. (e.g. time, log(time), or some step function of time)
  - A p value less than 0.05 gives statistical evidence that the HR varies with time
    - i.e. evidence of non-proportional hazards in that covariate
- Including allows us to relax the PH assumption yet still use Cox regression.

```
stcox arm cd4enrl, strata(stratavar) tvc(armn) texp( ln(_t))
```

Long format: Treat time as *discrete*

```
stset timefail indfail

stptime if ( armn==0 ), per(100)
stptime if ( armn==1 ), per(100)
```

Semi-parametric Cox model:

```
xi: stcox armn, hr
```

Proportional odds (logistic) model

The hazard is the probability that an event will occur in a given interval, given that the event has not already occurred (it is a conditional probability).

With continuous time, we can think of this as a slope.

With discrete time, we have the intervals and can calculate this directly.

Using either treatment of time, we can tie together hazard and survival.

```
logit _Y armn _d1-_d3, cluster(id) nocons or
```

Fit the proportional hazards (complimentary log-log) model

```
cloglog _Y armn _d1-_d3, cluster(id) nocons eform
```

```
sts graph, censored(number) failure by (armn) ///
        risktable(0 6.05 12.05 18.05 , failevents) ///
        ytitle (Cumulative Probability of Failure)  ///
        ylabel(0(.1)1.0) title(" ") ///
        xtitle("Months from HAART Initiation") ///
        xlabel(0(6)20) ///
        scheme(s2mono) ///
        plot1opts(recast(line) lcolor(black) lwidth(medthick) lpattern(dash)  connect(stairstep) cmissing(y)) ///
        plot2opts(recast(line) lcolor(black) lwidth(medthick) lpattern(solid) connect(stairstep) cmissing(y)) ///
        graphregion(fcolor(white) lcolor(white) ifcolor(white) ilcolor(white)) ///
        legend(off) ///
        tmax(18) ///
        text( 0.9 0 ///
                "Log-rank P-value<0.001" ///
                , place(se) box just(left) margin(l+4 t+1 b+1) width(77) bcolor(white) )
```

REFERENCES

Statistical analysis software (ex Stata, SAS)
http://www.ats.ucla.edu/stat/
http://www.cpc.unc.edu/research/tools/data_analysis/

Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls.
Pocock SJ, Clayton TC, Altman DG.
Lancet. 2002 May 11;359(9318):1686-9.

Stata presentation of survival data
From the help desk: Kaplan–Meier plots with stsatrisk
Jean Marie Linhart, Jeffrey S. Pitblado, James Hassell
The Stata Journal (2004) 4, Number 1, pp. 56–65

Comparison of Andersen-Gill and poisson and negative binomial regression
http://www.sciencedirect.com/science/article/pii/S0167947308002156

Stata analysis of recurrent events
http://www.stata.com/support/faqs/stat/stmfail.html

Longitudinal data analysis and survival data
http://data.princeton.edu/wws509/notes/c7.pdf

Discrete time survival analysis
http://gseacademic.harvard.edu/~willetjo/dsta.htm
Discrete-time survival analysis powerpoint slides
ALDA. Judith D. Singer & John B. Willett