'Recent' History of Survival Analysis in Clinical Trials

Dave Harrington dph@jimmy.harvard.edu

Harvard Statistics and Biostatistics, Dana-Farber Cancer Institute

12 August 2015

Some Context

Smith (1998)

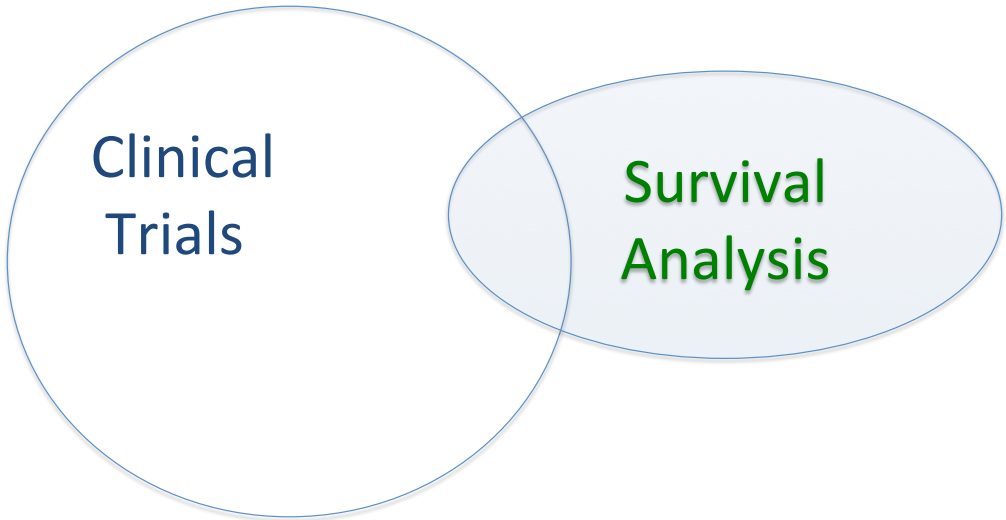
> *Was the randomised controlled trial the most important development in medicine this [20th] century?*

In T. Jefferson's 1807 letter to Caspar Washington [Peterson (1985)]

> *The patient, treated on the fashionable theory, sometimes gets well in spite of the medicine. The medicine therefore restored him, and the young doctor receives new courage to proceed in his bold experiments on the lives of his fellow creatures.*

All references at the end. First edition dates used for texts.

Why link clinical trials and survival analysis?



First ‘modern’ randomized clinical trial in 1948 [Marshall et al. (1948)]

Earliest work in estimating survival distributions seems to be Graunt’s 1661 life tables [Graunt (2008), Kindle (!) Edition].

Some important features of survival analysis responded to problems in clinical trials.

Clinical trials and the shift to non-parametric methods.

The study of reliability produced parametric methods for estimating life distributions. From Mann et al. (1974):

> *"the probability of a device (or item or organism) performing its (or his or her) defined purpose adequately for a specified period of time, under the operating conditions encountered"*

Early skepticism of modeling in JW Boag's discussion of Armitage (1959).

Difficult to determine when the use of non-parametric survival methods became routine for clinical trials.

Non-parametric approaches build momentum

Early non-parametric approaches:

- Berkson and Gage (1952) (estimation)
- Gehan (1965) (testing)

Non-parametrics 'flourished' with

- Kaplan and Meier (1958) (11^{th} most highly cited paper in Web of Science) and
- Mantel (1966), Peto and Peto (1972)

Elveback (1958) (118 citations) vs KM (approx 46,000 citations) shows power of marketplace: parametric vs. non-parametric.

More refined analysis of clinical trials

The early nonparametric approach did not support modeling, adjusting for confounders, etc.

- The proportional hazards model and its partial likelihood provided a modeling tool [Cox (1972), approx. 40,000 citations, 24^{th} most highly cited].
  - $\lambda(t|z) = \lambda_0(t)\exp(\beta' z)$
  - Contribution to the partial likelihood for β at failure time t_i

$$\frac{\exp(\beta' Z_{(i)})}{\sum_{l \in R_i} \exp(\beta' Z_l)}$$

- Early hints of the PH model in exponential regression [Feigl and Zelen (1965)].
- Accelerated failure time model provided an alternative.

More refined analysis of clinical trials

The early nonparametric approach did not support modeling, adjusting for confounders, etc.

- The proportional hazards model and its partial likelihood provided a modeling tool [Cox (1972), approx. 40,000 citations, 24^{th} most highly cited].
  - $\lambda(t|z) = \lambda_0(t)\exp(\beta' z)$
  - Contribution to the partial likelihood for β at failure time t_i

$$\frac{\exp(\beta' Z_{(i)})}{\sum_{l \in R_i} \exp(\beta' Z_l)}$$

- Early hints of the PH model in exponential regression [Feigl and Zelen (1965)].
- Accelerated failure time model provided an alternative.

Moving on from 1972

Many wonderful side trips . . .

- Buckley and James (1979) proposed using linear regression
- Competing risks, sample size calculations, model checking, interval censored data, examining efficiency, adjusting for tied failure times.

Kalbfleisch and Prentice (1980), Miller (1981), and Cox and Oakes (1984) provided important contributions and consolidations.

Some vexing problems in trials:

- Theory lacked the elegance of linear models
- Interim monitoring
- Causal inference in randomized trials with some non-adherence

Moving on from 1972

Many wonderful side trips . . .

- Buckley and James (1979) proposed using linear regression
- Competing risks, sample size calculations, model checking, interval censored data, examining efficiency, adjusting for tied failure times.

Kalbfleisch and Prentice (1980), Miller (1981), and Cox and Oakes (1984) provided important contributions and consolidations.

Some vexing problems in trials:

- Theory lacked the elegance of linear models
- Interim monitoring
- Causal inference in randomized trials with some non-adherence

First, the mathematics

Heroic efforts to pin down asymptotics using 'traditional methods'

- Kaplan-Meier product limit estimator: Breslow and Crowley (1974), others
- Proportional hazards model: Tsiatis (1981b), others
- Partial likelihood: Efron (1977), Wong (1986)

Structure came from surprising source: French school of probability [Meyer (1966)], via Aalen (1978). Partial likelihood score for β:

$$U(\beta, t) = \sum_i \int_0^t [Z_i(u) - E(\beta, u)] dM_i(u)$$

First, the mathematics

Heroic efforts to pin down asymptotics using 'traditional methods'

- Kaplan-Meier product limit estimator: Breslow and Crowley (1974), others
- Proportional hazards model: Tsiatis (1981b), others
- Partial likelihood: Efron (1977), Wong (1986)

Structure came from surprising source: French school of probability [Meyer (1966)], via Aalen (1978). Partial likelihood score for β:

$$U(\beta, t) = \sum_i \int_0^t [Z_i(u) - E(\beta, u)] dM_i(u)$$

The mathematics . . .

Important contributions using the martingale approach:

- The Kaplan-Meier estimate and the two sample problem: Gill (1980).
- The proportional hazards model: Andersen and Gill (1982)
- A general summary, with many applications Andersen et al. (1993). Affectionately, *ABGK*

Surprisingly useful in exploratory data analysis; Therneau and Grambsch (2000)

Made many difficult problems more accessible, e.g., sequential monitoring

Sequential Designs

Sequential monitoring was a 'not so immediate' response to the natural question: can we have a look at the data?

Some of the early work was not survival specific:

- Armitage (1960) important, early work (not survival focused).
- Haybittle (1971), early group sequential approach.
- Pocock (1977), the group sequential approach.
- O'Brien and Fleming (1979), for the boundaries most often used today.
- Lan and DeMets (1983), flexible computation of boundaries.

Tsiatis (1981a) used counting processes to show that all of the earlier technology could be applied in survival analysis, for commonly used tests.

Sequential Designs

Sequential monitoring was a 'not so immediate' response to the natural question: can we have a look at the data?

Some of the early work was not survival specific:

- Armitage (1960) important, early work (not survival focused).
- Haybittle (1971), early group sequential approach.
- Pocock (1977), the group sequential approach.
- O'Brien and Fleming (1979), for the boundaries most often used today.
- Lan and DeMets (1983), flexible computation of boundaries.

Tsiatis (1981a) used counting processes to show that all of the earlier technology could be applied in survival analysis, for commonly used tests.

Causal Inference

Methods for causal inference in observational studies well developed.

What do clinical trials add here? Immediacy, aversion to models . . .

- A clinical trial disrupted by non-adherence is distressing and expensive.
- Trialists studying chronic diseases are allergic to fully specified models.

Standard approach: Intent-to-Treat (ITT) in superiority RCTs, Per Protocol (PP) and ITT in non-inferiority RCTs.

Neither approach is very satisfying, despite FDA and EMA endorsement.

Causal inference . . .

Imbens and Rubin (2015) summarizes theory and applications of the *Rubin Causal Model*.

- Useful for binary or normally distributed 'immediate outcomes', not easy to adapt to censored event time data without a parametric model.

Robins and Tsiatis (1991) use Rank Preserving Structural Failure time models, or the structural version of the accelerated failure time model.

- Inference not based on assumptions about shape of baseline event rate function.
- Inspired by the prevalence of self-medication in the early days HIV clinical research.

Neither approach has substantially penetrated practice.

Causal inference . . .

Some side trips that show promise, but still used more often in observational studies

- Principal Stratification: Frangakis and Rubin (2002)
- Instrumental variables: Angrist et al. (1996)
- Propensity scores: Rosenbaum and Rubin (1983)

Emerging areas

- Adaptive designs. Tsiatis and Mehta (2003) and Gallo et al. (2006) give different views on this.
- Bayes. Slightly perverse to cite this as something that is coming. Spiegelhalter et al. (1994) very compelling argument for Bayesian designs and analysis. Ibrahim et al. (2005) consolidates what was known in 2005.
- Design and analyses for 'targeted' therapies biomarker adaptive designs [Jiang et al. (2007), Barker et al. (2009): I-SPY 2 trial].
- Patient-level prediction [Graf et al. (1999), Gerds et al. (2008)].

References

Aalen, O. (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics*, pages 701–726.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Science & Business Media.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics*, pages 1100–1120.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association*, 91(434):444–455.

Armitage, P. (1959). The comparison of survival curves. *Journal of the Royal Statistical Society. Series A (General)*, pages 279–300.

Armitage, P. (1960). Sequential medical trials. *Sequential Medical Trials.*

References (cont.)

Barker, A., Sigman, C., Kelloff, G., Hylton, N., Berry, D., and Esserman, L. (2009). I-spy 2: An adaptive breast cancer trial design in the setting of neoadjuvant chemotherapy. *Clinical Pharmacology & Therapeutics*, 86(1):97–100.

Berkson, J. and Gage, R. P. (1952). Survival curve for cancer patients following treatment. *Journal of the American Statistical Association*, 47(259):501–515.

Breslow, N. and Crowley, J. (1974). A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, 2(3):437–453.

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika*, 66(3):429–436.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 187–220.

Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*, volume 21. CRC Press.

References (cont.)

Efron, B. (1977). The efficiency of cox's likelihood function for censored data. *Journal of the American statistical Association*, 72(359):557–565.

Elveback, L. (1958). Estimation of survivorship in chronic disease: The âactuarialâ method. *Journal of the American Statistical Association*, 53(282):420–440.

Feigl, P. and Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information. *Biometrics*, pages 826–838.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

Gallo, P., Chuang-Stein, C., Dragalin, V., Gaydos, B., Krams, M., and Pinheiro, J. (2006). Adaptive designs in clinical drug development - an executive summary of the phrma working group. *Journal of biopharmaceutical statistics*, 16(3):275–283.

References (cont.)

Gehan, E. A. (1965). A generalized wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika*, 52(1-2):203–223.

Gerds, T. A., Cai, T., and Schumacher, M. (2008). The performance of risk prediction models. *Biometrical Journal*, 50(4):457–479.

Gill, R. D. (1980). Censoring and stochastic integrals. *Statistica Neerlandica*, 34(2):124–124.

Graf, E., Schmoor, C., Sauerbrei, W., and Schumacher, M. (1999). Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 18(17-18):2529–2545.

Graunt, J. (2008). *Natural and Political Observations made upon the Bills of Mortality*. Number 2. Evergreen Review, Inc.

Haybittle, J. (1971). Repeated assessment of results in clinical trials of cancer treatment. *The British journal of radiology*, 44(526):793–797.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2005). *Bayesian survival analysis*. Wiley Online Library.

References (cont.)

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

Jiang, W., Freidlin, B., and Simon, R. (2007). Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect. *Journal of the National Cancer Institute*, 99(13):1036–1043.

Kalbfleisch, J. D. and Prentice, R. L. (1980). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Kaplan, E. L. and Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481.

Lan, K. G. and DeMets, D. L. (1983). Discrete sequential boundaries for clinical trials. *Biometrika*, 70(3):659–663.

Mann, N. R., Schafer, R., and Singpurwalla, N. D. (1974). Methods for the statistical analysis of reliability and life data.

References (cont.)

Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer chemotherapy reports. Part 1*, 50(3):163–170.

Marshall, G., Blacklock, J., Cameron, C., Capon, N., Cruickshank, R., Gaddum, J., Heaf, F., Hill, A. B., Houghton, L., Hoyle, J. C., et al. (1948). Streptomycin treatment of pulmonary tuberculosis: a medical research council investigation. *BMJ*, 2(4582):769–782.

Meyer, P. A. (1966). *Probability and potentials*, volume 7. Blaisdell Publishing Company Waltham (Mass.).

Miller, R. G. (1981). *Survival analysis*, volume 66. John Wiley & Sons.

O'Brien, P. C. and Fleming, T. R. (1979). A multiple testing procedure for clinical trials. *Biometrics*, pages 549–556.

Peterson, M. (1985). Letters, 1760–1826. *Ed. Merrill D. Peterson. New York: Viking*.

References (cont.)

Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society. Series A (General)*, pages 185–207.

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2):191–199.

Robins, J. M. and Tsiatis, A. A. (1991). Correcting for non-compliance in randomized trials using rank preserving structural failure time models. *Communications in statistics-Theory and Methods*, 20(8):2609–2631.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Smith, R. (1998). Fifty years of randomized controlled trials. *BMJ*, 317:0.

References (cont.)

Spiegelhalter, D. J., Freedman, L. S., and Parmar, M. K. (1994). Bayesian approaches to randomized trials. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, pages 357–416.

Therneau, T. M. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer Science & Business Media.

Tsiatis, A. A. (1981a). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika*, 68(1):311–315.

Tsiatis, A. A. (1981b). A large sample study of cox's regression model. *The Annals of Statistics*, pages 93–108.

Tsiatis, A. A. and Mehta, C. (2003). On the inefficiency of the adaptive design for monitoring clinical trials. *Biometrika*, 90(2):367–378.

Wong, W. H. (1986). Theory of partial likelihood. *The Annals of Statistics*, pages 88–123.