The problem of attrition in longitudinal studies of the elderly

David Clayton

Emeritus Professor of Biostatistics, University of Cambridge, UK
Honorary Professor, London School of Hygiene and Tropical Medicine, UK

April 30, 2015

Acknowledgements

- Carole Dufouil
- The MELODEM (MEthods for LOngituinal studies in DEMentia) collaboration
- Niels Keiding

Outline

Attrition in studies of cognitive decline

In joint work in the 1990's (Dufouil et al., 2004) we noted that:

- most longitudinal data analyses did not differentiate between missing data due to dropout and missing data due to death,
- such analyses implicitly or explicitly impute cognitive function after death, while
- in the closely related area of competing risk analyses of disease incidence rates, analyses based on "death blocking" had fallen into disfavour.

We proposed a method which imputed the missing values for dropouts, but which did not so replace deaths. This talk is concerned with the treatment of deaths

On returning to the issue 15 years later . . .

- Our method has not gained widespread acceptance.
- It is argued that cognitive impairment is associated with increased mortality and, if the target of inference is an effect of exposure *specific* to cognitive function, this creates a “selection bias”.
- The controversy continues. An additional factor is the emergence of “causal models” — in particular the method of *principal stratification*

A recent exchange (*Epidemiology*, 2012)

- Weuve et al.(2012) estimated the effect of smoking on cognitive decline, dealing with attrition due to death by inverse-probability weighting (IPW)
- Chaix et al criticized this:
  "In replacing dead participants by cloning the living, IPW generates a sample in which participants are not allowed to die. . . . arguably a form of statistical cruelty"
- In a rejoinder, Tchetgen Tchetgen et al. (2012) noted
  "Although the cloning metaphor . . . can be a helpful pedagogic tool, it should not be taken too literally"
  arguing that IPW can alternatively be seen as a device to maintain independence between treatments and extraneous common causes of cognitive decline and mortality.

Outline

Diseases with a sharp onset; competing risks

There are two types of failure, here *death* or onset of *disease*. There have been two conceptual approaches to statistical analysis:

- **Latent failure times** Denoting the *potential* times of disease onset and death by T and S, we can observe only Min(T, S)
- **Cause-specific hazard rates** The instantaneous rates of occurrence of disease onset and death conditional upon avoidance of *both* causes. The former is known to epidemiologists as the disease *incidence rate*.

The question for aetiological research: If an exposure affects both disease onset and death, can we measure a *specific* effect on disease?

Potential failure times

Apart from any philosophical doubts:

- For the specific effect question, we are interested in the distribution of disease onset times conditional on exposure(s), $\Pr(T|X)$.
- But, if exposure also affects time of death, S, this depends on the joint distribution: $\Pr(T|X) = \sum_S \Pr(T, S|X)$.
- In the competing risks framework, times of disease onset and death cannot both be observed in the same individual, so that the joint distribution cannot be studied empirically.
- Without untestable assumptions, usually *independence* of S and T given X, $T \perp\!\!\!\perp S|X$, the model is uninterpretable
- There are bounds on specific effects, but usually these are usually very broad

Cause-specific hazards and "death–blocking"

Not cause-specific! The incidence rate of a specific disease is defined with respect to a population who have survived *all* causes. One approach to measuring specific effects is **death–blocking**

- Estimate effects on the probability of developing disease in a specified period *if there was no mortality*
- Aside from the obvious difficulty in envisaging such a scenario, this also requires the conditional independence assumption — conditional on exposure(s), incidence rates for the disease would be unchanged by preventing deaths

! Without this assumption, the approach is still defensible, providing a view of the *bivariate* response, via the factorization

$$\Pr(\text{Survival, Disease}) = \Pr(\text{Survival}).\Pr(\text{Disease}|\text{Survival})$$

Historical note: the elimination of smallpox

- **Bernoulli, D'Alembert, Laplace (18^{th} century)**
  - All offered solutions to the question "What is the probable effect upon the increase of population of the extinction of smallpox?"
- **Makeham (1874)**: "On the theory of the Composition of Decremental Forces"
  - "It will be observed that these solutions all proceed on the assumption that the extermination of small pox does not affect the mortality arising from other causes. This must be proved before any reliance can be placed upon the conclusions arrived at"
  - But he did not discuss how such proof could be obtained

More recently

- **Cornfield (1957)**: "The Estimation of the Probability of Developing a Disease in the Presence of Competing Risks"
  - "The reader who is — shall we say — not comfortably at home in mathematics, the appearance of the illustrative computation in this discussion may tempt him to turn to the next paper. No matter what his interests, one cannot help but gain from this presentation a wholesome caution about accepting seemingly obvious conclusions"
  - "The question 'What is the probability of developing a particular disease?' is not an unambiguous one"

Semicompeting risks: the illness–death model

Recent work which purports to break the impasse is based on the fact that times of death and disease onset *can* both be observed — if disease occurs first (Varadhan et al., 2014).

- **Latent failure times**: extrapolate from the joint distribution of $(T, S; S > T)$ to the whole distribution using copula models
- **Hazard rates**: In the illness–death model use random effect (shared "frailty") models, extrapolating from
  - (identifiable) shared frailty effects between Health$\rightarrow$Disease and Disease$\rightarrow$Death transitions, to
  - (non-identifiable) shared frailty between Health$\rightarrow$Disease and Health $\rightarrow$Death transitions

In most areas of epidemiology, such extrapolation would not be widely regarded as legitimate

Cumulative incidence

There has been renewed interest in this measure, defined as the probability of contracting disease before death (or, if surviving, before a defined truncation time)

- Not attractive for studies of disease aetiology as it is too dependent on effects on mortality
- Although arguably of relevance to public health and pragmatic analyses of clinical trials, this measure would need to be accompanied by estimates of the *duration* of disease
- This can be seen as an alternative view of the bivariate outcome, based on the factorization

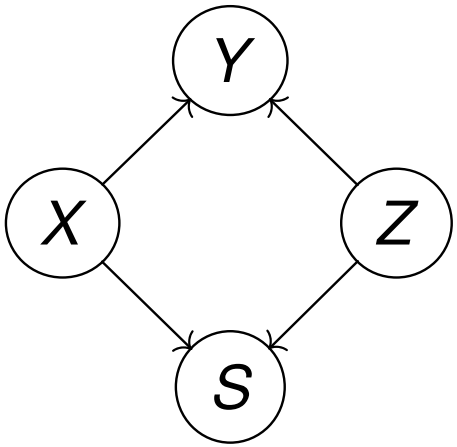
$$\Pr(\text{Survival, Disease}) = \Pr(\text{Disease}).\Pr(\text{Survival}|\text{Disease})$$

Outline

Cognitive decline

- A closely related problem: we have longitudinal measurements of cognitive function (e.g. MMSE), and treatments/exposures which may affect both cognition and mortality (Kurland et al., 2009)
- A widely expressed aspiration: to estimate effects of treatment/exposure *specifically on cognitive function*
- In assessing methods which purport to achieve this, we must ask
  - Are the necessary assumptions plausible?
  - Even if they are, are the effect estimates meaningful?
- Many of the issues can be demonstrated considering only the first follow–up measurement in a randomized study

A randomized trial



- Y: cognitive function at first follow–up
- S: survival until first follow–up
- X: a randomized intervention
- Z: a "common cause" — an aetiological factor which affects both cognitive function and survival

- Because X is assigned randomly, $Y \perp\!\!\!\perp S|X, Z$ so that, at randomization, Z is not a confounder for the effect of X on Y
- But, conditioning on survival ($S = 1$) **as we must**, induces an association between X and Z
- This introduces what has been termed *collider-stratification bias* or *selection bias* (Hernan et al., 2004)

A numerical example

Risk stratum Z	Exposed, $X = 1$			Unexposed, $X = 0$		
	N	n	$\overline{Y}$	N	n	$\overline{Y}$
Low, $Z = 1$	100	80	28	100	90	30
Medium, $Z = 2$	100	60	25	100	80	28
High, $Z = 3$	100	40	20	100	70	24
Marginal	300	200	25.22	300	240	27.58

- N is the number of subjects in each category at randomization, and n is the corresponding number surviving. $\overline{Y}$ is the mean cognitive function score
- Mean cognitive function, $\overline{Y}$, is lower in exposed group, and mortality is higher. Mean cognitive score and probability of survival both decrease with increasing Z
- The marginal effect in survivors, $25.22 - 27.58 = -2.36$, is confounded as a result of differential survival

A comment on the language: bias or confounding?

- Describing this as *selection bias* is at odds with earlier usage, in which "bias" is introduced by inadequacies of design or analysis
- For example (Breslow and Day, Vol 1, 1980, p85):
  "Confounding reflects the causal association between variables in the population under study. Bias, by contrast, is not a property of the underlying population . . . It results from inadequacies of the design"
  Measuring cognitive function only in survivors can hardly be regarded as an inadequacy of the design!
- If the inferential targets are effects specific to cognitive function, the inevitable conditioning on survival creates widespread *confounding*

Deaths as missing data: standardization

Risk stratum Z	Exposed, $X = 1$			Unexposed, $X = 0$		
	N	$n^{[DB]}$	$\overline{Y}$	N	$n^{[DB]}$	$\overline{Y}$
Low, $Z = 1$	100	100	28	100	100	30
Medium, $Z = 2$	100	100	25	100	100	28
High, $Z = 3$	100	100	20	100	100	24
Marginal	300	300	24.33	300	300	27.33

- Weighting by the inverse of the probability of death restores the survivor population to its initial structure. This is *direct standardization* for Z — using the initial population distribution of Z as the standard population. It results in a larger estimated effect, -3.00 rather than -2.36
- This illustrates the remark of Tchetgen Tchetgen — that the cloning metaphor should not be taken too literally — it can be regarded simply as a device to remove the confounding

Is this an appropriate standard population?

- Direct standardization can be viewed as simulating a randomized experiment. But does this weighting simulate an experiment we could imagine?
  - The hypothetical study population is immortal
  - As a consequence, some groups with high mortality become heavily up–weighted, leading to unstable estimates
- There are many alternative weightings which achieve standardization — we just require

$$\frac{\text{Weight}(X=1,Z)}{\text{Weight}(X=0,Z)} = \frac{\Pr(\text{Survival}|X=0,Z)}{\Pr(\text{Survival}|X=1,Z)}.$$

Different choices lead to different standard populations

"Causal" models: what are they?

- With some additional assumptions and used properly (e.g. not conditioning on intermediate variables) almost any probability model can be a causal model
- Models described explicitly as causal have deterministic elements
  - Structural equation models
  - Potential responses ("counterfactuals")
- These are often *marginal* models
  - Effect estimates are averaged over the distribution of concomitant variables, rather than conditional upon fixing their values
  - They often have strong connections with direct standardization

Principal stratification (Rubin, 2000)

- Posit the existance of *potential responses* — the values of cognitive score and survival when $X = 1$ and when $X = 0$, denoted by (Y_1, S_1) and (Y_0, S_0). But we only observe one of these pairs in a single subject
- The *survivor average causal effect* (SACE) is defined as

$$\mathrm{E}(Y_1 - Y_0 | S_1 = 1, S_0 = 1)$$

  — the average difference between exposed and unexposed potential responses in subjects who would have survived regardless of exposure
- This is not the same as the (observable) crude effect $\mathrm{E}(Y_1 | S_1 = 1) - \mathrm{E}(Y_0 | S_0 = 0)$, which is confounded
- Because we cannot observe both pairs of potential responses in one subject, we must make assumptions about their interrelationship

Explainable nonrandom survival[1]

- Assumes, conditional on covariates, substantial independence between exposed and unexposed potential responses. Formally:

$$S_1 \perp\!\!\!\perp S_0|Z, \qquad S_1 \perp\!\!\!\perp Y_0|Z, S_0 = 1, \qquad S_0 \perp\!\!\!\perp Y_1|Z, S_1 = 1$$

- Hayden et al. show, in the case of a randomized study, that the SACE can be estimated by the difference in mean scores when exposed subjects are weighted by $\Pr(S = 1|X = 0, Z)$ and unexposed subjects are weighted by $\Pr(S = 1|X = 1, Z)$
- These weights conform with the conditions to provide standardization for confounding by Z created by differential survival
- The corresponding standard population is that expected under a matched pair study — when both members of a pair are removed from study on the death of either one

[1] Hayden et al.,2004 following Robins, 1998

Our numerical example

Risk stratum Z	Exposed, $X = 1$ N	$n^{[ENS]}$	$\overline{Y}$	Unexposed, $X = 0$ N	$n^{[ENS]}$	$\overline{Y}$
Low, $Z = 1$	100	72	28	100	72	30
Medium, $Z = 2$	100	48	25	100	48	28
High, $Z = 3$	100	28	20	100	28	24
Marginal	300	148	25.51	300	148	28.22

- $n^{[ENS]}$ is the expected number of subjects remaining in a matched pair study. E.g. for $Z = 1$ the survival probabilities were .8 and .9 for exposed and unexposed subjects, so that $n^{[ENS]} = 100 \times .8 \times .9 = 72$
- The estimated SACE is $25.51 - 28.22 = -2.71$, intermediate between the crude effect and "death–blocking" estimates
- But this standard population has unrealistically high mortality

Monotone survival

- Tchetgen Tchetgen (2013), taking smoking as the exposure of interest, assumed that a subject who would survive when exposed would also survive when not exposed
- The main estimate of the SACE proposed was the difference between weighted mean cognitive function scores, where

$$\begin{aligned} \text{Weight} &= \frac{\Pr(\text{Survival}|X=1,Z)}{\Pr(\text{Survival}|X=0,Z)}, && \text{if } X=0 \\ &= 1 && \text{if } X=1 \end{aligned}$$

- Again, in our randomized setting, these weights conform with the conditions to standardize for confounding by Z in survivors
- The corresponding standard population is one in which both exposed and unexposed groups suffer the same mortality as was observed in the exposed (high mortality) group

Our numerical example

Risk stratum Z	Exposed, $X = 1$			Unexposed, $X = 0$		
	N	$n^{[MS]}$	$\overline{Y}$	N	$n^{[MS]}$	$\overline{Y}$
Low, $Z = 1$	100	80	28	100	80	30
Medium, $Z = 2$	100	60	25	100	60	28
High, $Z = 3$	100	40	20	100	40	24
Marginal	300	180	25.22	300	180	28.00

- $n^{[MS]}$ is the expected number of subjects surviving if mortality in both groups were the same as observed in the exposed group
- The estimate of the SACE is $25.22 - 28.00 = -2.78$, somewhat larger than that obtained under the explainable nonrandom survival assumption

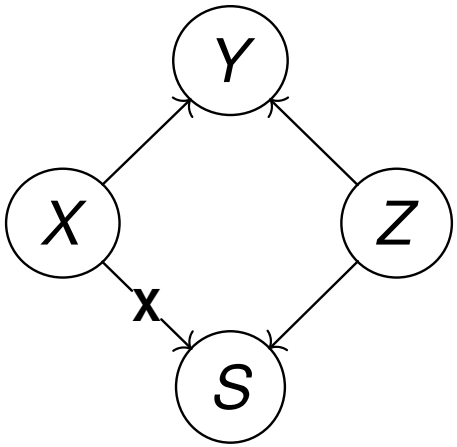
Conditional (regression) models

- In this case, treating deaths as missing data and the principal stratification approaches are equivalent to direct standardization
- They estimate marginal effects in standard survivor populations subject to different assumptions concerning mortality — none, in my view, fully satisfactory
- An alternative, conditional, approach is to fit the regression model in survivors

$$\mathrm{E}(Y|X, Z, S = 1) = \alpha + \beta X + \gamma z$$

- If no interaction, β should be equal to the standardized marginal effects — we would get the same answer by any of the methods

Summary — as a graph



No further unmeasured common causes of Y and S

- Death blocking and principal stratification approaches block the path $X \rightarrow S$ by (counterfactually) making mortality independent of exposure
- Mechanism?
- Regression model for cognitive score, Y, conditions on X and Z so that conditioning on survival has no effect

Difficulties with "causal modelling" approaches

- Principal stratification makes strong assumptions (most importantly that of *no unmeasured common aetiology*) and the interpretation of estimates is not clear when these assumptions are not met
- The interpretation in terms of standardization goes some way to correcting this, but what standard population should be used?
  - Can such a population be relevant to real life, rather than rather artificial notional experiments?
  - Is there a plausible *mechanism* which would create the such a counterfactual population?
  - E.g. would we really wish to consider a mechanism which rendered mortality the same in non-smokers as in smokers?

Regression analysis in survivors

- The alternative is to accept that the response is bivariate and analyse the observed (survivor) data using the factorization

$$\Pr(Y, S = 1|X, Z) = \Pr(Y|X, Z, S = 1)\Pr(S = 1|X, Z)$$

e.g.

  - regression models for Y against X and Z in survivors
  - (logistic) regression models for survival against X and Z

- When interpreting the former regression, we need to be aware of the fact that effects of exposure are potentially influenced by differential survival — as when interpreting incidence rates
- In the absence of interaction, this approach would give exactly the same estimate of exposure effect as principal stratification and death–blocking

Outline

Observational studies

Now consider the case of an observational study in which Z is a confounder even before any deaths have occurred:

- The weighting SACE methods so far considered no longer coincide with direct standardization
- Tchetgen Tchetgen suggested multiplying the weights by a further correction — the *propensity score* for exposure:

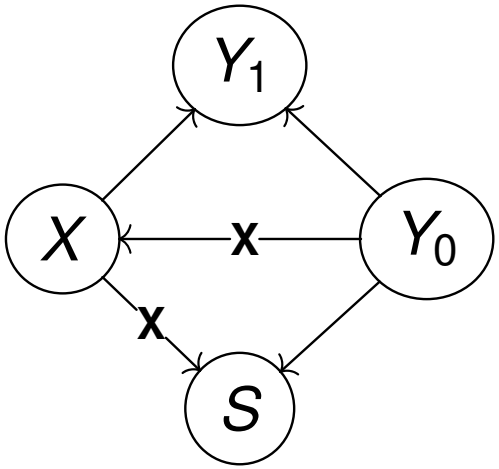
$$\{\Pr(X|Z)\}^{-1}$$

- The method then returns to being identical to direct standardization — now correcting for the *total* confounding by Z. The two components of the weights correct the two sources of confounding
  - that due to the initial association between Z and exposure
  - the additional confounding due to differential survival.

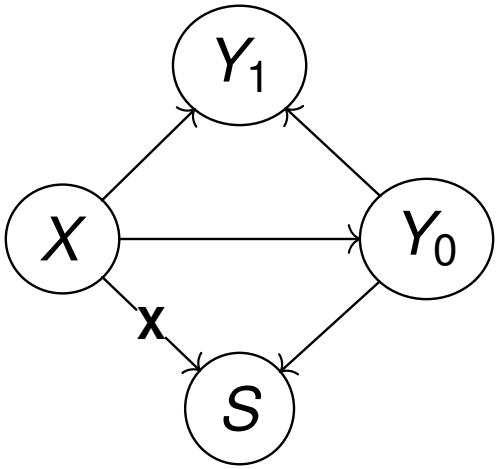
Intermediate variables

- In the first sections of his paper, Tchetgen Tchetgen allowed for an initial relationship between X and Z, but did not correct for this
- This is because he took Z to be an earlier measurement of Y, made after assignment of exposure — it is an *intermediate variable*.
- In the case where Z is a pre-exposure measurement, it is appropriate to standardize for both components of confounding by Z, while
- where Z is a post-exposure measurement it is appropriate to standardize only for the differential mortality component

As graphs . . .



(a) Y_0 precedes exposure



(b) Exposure precedes Y_0

- In case (a), the two sets of weights block $X \rightarrow S$ and $Y_0 \rightarrow X$
- In case (b) we do not block $X \rightarrow Y_0$ as it is on the causal path
- Note that the regression modelling approach to case (b) is an autoregressive (AR) model. A problem is that no single parameter represents the total effect of X on Y_1

Longitudinal studies with multiple waves

- Principal stratification
  - Unlike most discussions of principal stratification, Tchetgen Tchetgen extended his method to multiple waves. Survival across waves is modelled using life table methods and weights calculated using fitted survival probabilities
  - He treated the first measure as case (a) and subsequent measures as case (b). (But does the smoking example fit?)
- Modelling longitudinal responses of survivors
  - Our earlier work considered, mainly, GEE models — equivalent to analysis of *prevalence*
  - We also indicated that analysis which focused on *changes* in cognitive function would be more in the spirit of analyses of *incidence* — perhaps AR models are a possibility?
  - As elsewhere, endogenous covariates present difficulties

Final remarks

- Causal models can be “good ideas taken too far”
- Cognitive decline and death should be treated as a bivariate outcome
- Parametric models aren’t such a bad thing

Causal models can be "good ideas taken too far"[1]

- Although the recent careful discussion of causality has been helpful, there is a negative side
- There are strong assumptions and the language invites overambitious conclusions
- Here, the assumption of no unmeasured shared aetiological factors is unlikely to be true
- What do "average causal effects" mean when we are not confident of the assumptions?

[1] Joffe (2011)

Cognitive decline and death should be treated as a bivariate outcome

- If the assumption of no unmeasured shared aetiological factors is unsupportable, and if an exposure affects both mortality and cognitive function, we will not be able to partition this into separate causal effects. We must regard the response as bivariate
- The factorization

$$\Pr(Y, S = 1|X, Z) = \Pr(Y|X, Z, S = 1)\Pr(S = 1|X, Z)$$

is natural, and has an interpretation when the assumptions necessary to compute a SACE are not met. But there are other possibilities

Parametric models aren't such a bad thing

- Much is made of the fact that average causal effect estimates are based on a nonparametric approach
- But, if the effect is heterogeneous, the average effect will depend on the (counterfactual) population over which it is averaged. The nature of that population and the mechanism under which it could arise are often vague
- Classical parametric approaches, such as regression models, when properly used, can be causal models
- These can (and should) explore effect heterogeneity, potentially leading to improved understanding

N. Breslow and N. Day. *Statistical Methods in Cancer Research. Volume I — The Analysis of Case–Control Studies*. IARC Scientific Publications. IARC, Lyon, 1980.

B. Chaix, D. Evans, J. Merlo, and E. Suzuki. Weighing up the dead and missing: Reflections on inverse probability weighting and principal stratification by death. *Epidemiology*, 23, 2012.

C. Dufouil, C. Brayne, and D. Clayton. Analysis of longitudinal studies with death and dropout: A case study. *Stat. Med.*, 23:2215–2226, 2004.

D. Hayden, D. Pauler, and D. Schoenfeld. An estimator for treatment comparisons among survivors in randomized trials. *Biometrics*, 61:305–310, 2005.

M. Hernán, S. Hernádez-Díaz, and J. Robins. A structural approach to selection bias. *Epidemiology*, 15:615–625, 2004.

M. Joffe. Principal stratification and attribution prohibition: good ideas taken too far. *Int. J. Biostat.*, 7(1), 2011.

B. F. Kurland, L. L. Johnson, B. L. Egleston, and P. H. Diehr. Longitudinal data with follow-up truncated by death: match the analysis method to research aims. *Stat. Sci.*, 24:211–222, 2009.

J. Robins. Correction for non-compliance in equivalence trials. *Stat. Med.*, 17:269–302, 1998.

D. Rubin. Comment on “Causal inference without counterfactuals” by A.P. Dawid. *J. Am. Stat. Ass.*, 95:435–438, 2000.

E. Tchetgen Tchetgen. Identification and estimation of survivor average causal effects. *Stat. Med.*, 33:3601–3628, 2013.

E. Tchetgen Tchetgen, M. Glymour, I. Shpitser, and J. Weuve. To weight or not to weight? on the relation between inverse probability weighting and principal stratification for truncation by death. *Epidemiology*, 23:132–137, 2012.

R. Varadhan, Q. Li Xue, and K. Bandeen-Roche. Semicompeting risks in aging research: methods, issues, and needs. *Lifetime Data Anal.*, 20(4):538–562, 2014.

J. Weuve, E. Tchetgen Tchetgen, M. Glymour, et al. Accounting for bias due to selective attrition: the example of smoking and cognitive decline in older adults. *Epidemiology*, 23:119–128, 2012.