Longitudinal Data with Follow-up Truncated by Death: Matching the Analysis Method to Research Aims

Brenda F. Kurland
Fred Hutchinson Cancer Research Center

Objectives

- Censoring due to death is different from data missing due to dropout
- Analysis methods for longitudinal data truncated by death
  - Targets of inference dictated by factorization of joint distribution of longitudinal response (Y) and survival (S)
  - Choose analysis method to address research aims
Data missing due to death: The Cardiovascular Health Study (CHS)

- Population-based prospective longitudinal study, n=5888 age 65+ at baseline
- 3MSE (0-100) for cognitive functioning, annually for up to 10 years
- We examine n=3814 age 70+ at baseline
  - 44% of men die during follow-up
  - 29% of women die during follow-up
- Questions of interest
  - What is the rate of change over time in 3MSE?
  - What is the expected cognitive status at a given age?
## Hypothetical CHS 3MSE data

<table>
<thead>
<tr>
<th></th>
<th>Age 70</th>
<th>Age 71</th>
<th>Age 72</th>
<th>Age 73</th>
<th>Age 74</th>
<th>Age 75</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1 “normal”</strong></td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td><strong>2 “mild cognitive impairment”</strong></td>
<td>84</td>
<td>82</td>
<td>80</td>
<td>78</td>
<td>76</td>
<td>74</td>
</tr>
<tr>
<td><strong>3 “terminal decline”</strong></td>
<td>84</td>
<td>80</td>
<td>76</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><strong>4 “terminal decline”</strong></td>
<td>65</td>
<td>50</td>
<td>35</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
### Statistical model #1: Joint model $f(Y_i, S_i)$, percent alive and healthy

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70</td>
<td>71</td>
<td>72</td>
<td>73</td>
<td>74</td>
<td>75</td>
</tr>
<tr>
<td>1 “normal”</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>2 “mild cognitive impairment”</td>
<td>84</td>
<td>82</td>
<td>80</td>
<td>78</td>
<td>76</td>
<td>74</td>
</tr>
<tr>
<td>3 “terminal decline”</td>
<td>84</td>
<td>80</td>
<td>76</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4 “terminal decline”</td>
<td>65</td>
<td>50</td>
<td>35</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

- Define “healthy” as 3MSE ≥ 80
- 75% alive and healthy at age 70
- 25% alive and healthy at age 75
Statistical model #1: Joint model $f(Y_i, S_i)$, percent alive and healthy

- Define “healthy” as $3MSE \geq 80$
- Empirical percent alive and healthy shown: may also model
- Not monotone: possible to regain health
- Describes entire cohort at each timepoint

Analysis method: logistic regression or GEE
Diehr et al., *Medical Care* 1995; Diehr et al., *J Clin Epi* 1998
Statistical model #1: Joint model f(Yi, Si), percent alive and healthy

- Percent alive and healthy at age 75
  - 0.82 (females)
  - 0.75 (males)
- Percent alive and healthy at age 79
  - 0.70 (females)
  - 0.54 (males)
- Joint model: reflects both survival and 3MSE (but doesn’t separate)

Analysis method: logistic regression or GEE

Diehr et al., *Medical Care* 1995; Diehr et al., *J Clin Epi* 1998
### Statistical model #2: Unconditional f(Y_i)

<table>
<thead>
<tr>
<th>Age</th>
<th>70</th>
<th>71</th>
<th>72</th>
<th>73</th>
<th>74</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 &quot;normal&quot;</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>2 &quot;mild cognitive impairment&quot;</td>
<td>84</td>
<td>82</td>
<td>80</td>
<td>78</td>
<td>76</td>
<td>74</td>
</tr>
<tr>
<td>3 &quot;terminal decline&quot;</td>
<td>84</td>
<td>80</td>
<td>76</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4 &quot;terminal decline&quot;</td>
<td>65</td>
<td>50</td>
<td>35</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

- **Expected 3MSE at age 75:**

  $E(Y|\text{age}=75) = \frac{E(Y|\text{age}=75, S>75) \cdot P(S>75)}{2} + \frac{E(Y|\text{age}=75, S \leq 75) \cdot P(S \leq 75)}{2}$

  $= \frac{54 + 35}{2} \cdot \frac{1}{2} + \frac{35 + 0}{2} \cdot \frac{1}{2}$

  $= \frac{89}{2} \cdot \frac{1}{2} + \frac{35}{2} \cdot \frac{1}{2}$

  $= \frac{44.5}{2} + \frac{17.5}{2}$

  $= 22.25 + 8.75 = 54.5$
Statistical model #2: Unconditional $f(Y_i)$

- quadratic in time
- random intercept and 1st degree polynomial
- individual fitted trajectories resemble observed trajectories
- time trends of mean model implicitly impute beyond death

Analysis method: linear mixed model (aka random effects, hierarchical linear)

Laird & Ware, *Biometrics* 1982
Statistical model #2: Unconditional f(Y_i)

- Expected 3MSE for both males and females
  - 86 at age 75
  - 77 at age 79
- strong age-associated decline in cognitive functioning

Analysis method: linear mixed model (aka random effects, hierarchical linear)
Laird & Ware, *Biometrics* 1982
### Statistical model #3: Fully conditional \( f(Y_i \mid S_i=s) \)

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70</td>
</tr>
<tr>
<td>1 &quot;normal&quot;</td>
<td>90</td>
</tr>
<tr>
<td>2 &quot;mild cognitive impairment&quot;</td>
<td>84</td>
</tr>
<tr>
<td>3 &quot;terminal decline&quot;</td>
<td>84</td>
</tr>
<tr>
<td>4 &quot;terminal decline&quot;</td>
<td>65</td>
</tr>
</tbody>
</table>

- different slopes for two groups
- different expected values for two groups
Statistical model #3: Fully conditional $f(Y_i \mid S_i=s)$

- quadratic in time
- stratified by year of death relative to baseline
- terminal decline in decedents
- relatively stable functioning in survivors

Analysis method: pattern mixture model (fitted using random effects model)
Little, JASA 1995; Pauler et al., Stat Med 2003
Statistical model #3: Fully conditional \( f(Y_i | S_i=s) \)

- Expected 3MSE
  - 75–91 at age 75, depending on stratum
  - 84 (female), 86 (male) at age 79 (survivors only)

- Fitted average trajectories condition on survival time, which is not known prospectively

Analysis method: pattern mixture model (fitted using random effects model)
Little, JASA 1995; Pauler et al., Stat Med 2003
### Statistical model #4: Fully conditional \( f(Y_i \mid S_i=s) \), terminal decline

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>70</td>
</tr>
<tr>
<td>1 “normal”</td>
<td>90</td>
</tr>
<tr>
<td>2 “mild cognitive impairment”</td>
<td>84</td>
</tr>
<tr>
<td>3 “terminal decline”</td>
<td>84</td>
</tr>
<tr>
<td>4 “terminal decline”</td>
<td>65</td>
</tr>
</tbody>
</table>

- decedents only
- slope for years approaching death
Statistical model #4: Fully conditional $f(Y_i | S_i=s)$, terminal decline

- same linear mixed model as for #1, but only decedents and time scale is years to death
- about 5 pt/year decline
- expected average 3MSE
  - 85 at age 75
  - 82 at age 79

Analysis method: random effects model
Siegler, *Experimental Aging Research* 1975;
Diehr et al., *J Clin Epi* 2002; Wilson et al., *Neurology* 2003
### Statistical model #5: Partly conditional $f(Y_i \mid S_i > t)$, regression conditioning on being alive

<table>
<thead>
<tr>
<th>Age</th>
<th>70</th>
<th>71</th>
<th>72</th>
<th>73</th>
<th>74</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 “normal”</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>2 “mild cognitive impairment”</td>
<td>84</td>
<td>82</td>
<td>80</td>
<td>78</td>
<td>76</td>
<td>74</td>
</tr>
<tr>
<td>3 “terminal decline”</td>
<td>84</td>
<td>80</td>
<td>76</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>4 “terminal decline”</td>
<td>65</td>
<td>50</td>
<td>35</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>average</td>
<td>81</td>
<td>76</td>
<td>70</td>
<td>84</td>
<td>83</td>
<td>82</td>
</tr>
</tbody>
</table>

- prevalent 3MSE for survivors
Statistical model #5: Partly conditional $f(Y_i | S_i > t)$, regression conditioning on being alive

- Independence correlation structure avoids implicit imputation
- Average 3MSE declines 1 pt/year: describes dynamic cohort of survivors, not trajectory for individuals
- Average prevalent 3MSE
  - 91 at age 75
  - 87 at age 79

Analysis method: GEE with independence working correlation
Kurland & Heagerty, *Biostatistics* 2005
Statistical model #6: Fully conditional 
\[ f(Y_i \mid S_i(0)>s, S_i(1)>s) \] principal stratification

- Men are generally at greater risk of death
- Do women suffer lower physical and mental decline, or do they survive with greater deficits than men?
- Causal effects for selected principal strata defined by potential survival outcomes

Analysis method: principal stratification model
**Statistical model #6: Fully conditional**
\[ f(Y_i | S_i(0)>s, S_i(1)>s) \]  principal stratification

- Gender and cognitive function trajectory in stratum of patients expected to live 9 years regardless of gender

- Men and women both stable for 4 years, decline a few points over next 5 years

Analysis method: principal stratification model
Summary

- Implicit imputation will occur if random effects models applied to unbalanced data.
- Prevalent response values over time may not represent trajectories of individuals.
Conclusions

- Choice of analysis has a great influence on interpretation of longitudinal data truncated by death
- Separate decision process for target of inference and analysis method
  - Marginalized models

Shameless plug (again)

Acknowledgments

- University of Washington
  - Patrick Heagerty
  - Thomas Lumley

- Cardiovascular Health Study
  - NHLBI: N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01 HC-15103, N01 HC-55222, N01-HC-75150, N01-HC-45133, U01 HL080295
  - Intramural Research Program of the NIH
  - P30 CA 06927
  - Commonwealth of Pennsylvania
  - Additional contribution from the NINDS
  - A full list of principal CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm.
Discussion: Analysis selection

- Health care providers planning demand for dementia services
- Randomized clinical trial of supportive care in hospice patients
- Individual differences in cognitive decline