IOWA HEALTH PROFESSIONS INVENTORY
— Integrated Tracking —
IOWA PHYSICIAN INFORMATION SYSTEM

• Continuous inventory
• All Iowa physicians (MD/DO)
• Purposes:
  – characterize physician population
  – monitor trends
  – support and evaluate programs
  – conduct research
  – inform policy
  – produce data products/services
DATA FIELDS

• Name and practice address
• Degree (MD/DO)
• Unique ID numbers (3)
• Gender
• Birthdate
• Birthstate
• Medical college
• Specialty/subspecialty
• Residency/fellowship training
DATA FIELDS (cont.)

- Certification and recertification
- Professional activity
- Practice arrangement/relationships
- Worksites (principal/VCCs/satellites)
- History (temporal/dates of events)
- Clinical teaching/appointments
- Demographics: community/county
Why do doctors leave Iowa?

Can we actively change attrition?
(number of personnel leaving)

• Can we find a ‘determining’ characteristic related to attrition?
• Maybe work environment?
• Or another controllable characteristic?
## View of data

<table>
<thead>
<tr>
<th>Degree</th>
<th>Gender</th>
<th>BirthYear</th>
<th>Worksite</th>
<th>Effect year</th>
<th>Action</th>
<th>Citypop</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO</td>
<td>M</td>
<td>1960</td>
<td>Knoxville Hospital Clinic</td>
<td>1995</td>
<td>ADD</td>
<td>7731</td>
</tr>
<tr>
<td>DO</td>
<td>M</td>
<td>1960</td>
<td>Iowa Lutheran Hospital</td>
<td>1997</td>
<td>TRF</td>
<td>198682</td>
</tr>
<tr>
<td>MD</td>
<td>F</td>
<td>1959</td>
<td>Dept of Anesthesia</td>
<td>1996</td>
<td>ADD</td>
<td>62220</td>
</tr>
<tr>
<td>MD</td>
<td>F</td>
<td>1959</td>
<td>Linn County Anesthesiologists</td>
<td>1998</td>
<td>TRF</td>
<td>6480</td>
</tr>
<tr>
<td>MD</td>
<td>M</td>
<td>1965</td>
<td>Great River Medical Center</td>
<td>1995</td>
<td>ADD</td>
<td>3161</td>
</tr>
<tr>
<td>MD</td>
<td>M</td>
<td>1965</td>
<td>Great River Medical Center</td>
<td>2002</td>
<td>DEL</td>
<td>3161</td>
</tr>
<tr>
<td>MD</td>
<td>M</td>
<td>1959</td>
<td>Dubuque Urology Services</td>
<td>1995</td>
<td>ADD</td>
<td>57686</td>
</tr>
<tr>
<td>MD</td>
<td>M</td>
<td>1959</td>
<td>Dubuque Urology Services</td>
<td>2004</td>
<td>DEL</td>
<td>57686</td>
</tr>
<tr>
<td>MD</td>
<td>F</td>
<td>1964</td>
<td>Alegent Hlth Missouri ValleyCl</td>
<td>1996</td>
<td>ADD</td>
<td>2992</td>
</tr>
<tr>
<td>MD</td>
<td>F</td>
<td>1964</td>
<td>Alegent Hlth Missouri ValleyCl</td>
<td>2001</td>
<td>DEL</td>
<td>2992</td>
</tr>
<tr>
<td>MD</td>
<td>M</td>
<td>1961</td>
<td>VA Central Iowa Hlth Care Syst</td>
<td>1995</td>
<td>ADD</td>
<td>7731</td>
</tr>
<tr>
<td>MD</td>
<td>M</td>
<td>1938</td>
<td>Finley Neurological Services</td>
<td>1995</td>
<td>ADD</td>
<td>57686</td>
</tr>
<tr>
<td>MD</td>
<td>M</td>
<td>1938</td>
<td>Finley Neurological Services</td>
<td>2001</td>
<td>DEL</td>
<td>57686</td>
</tr>
<tr>
<td>MD</td>
<td>M</td>
<td>1966</td>
<td>McFarland Clinic PC</td>
<td>1996</td>
<td>ADD</td>
<td>50731</td>
</tr>
</tbody>
</table>
How do we quantify attrition in this data set? (or predict it)

• If it was a cross-section (i.e. single point in time), could think of logistic regression.

• Since we have information across time, perhaps we could think of this as a survival analysis…
Survival analysis

• Studies time to event (often, death)

• Subjects may not all join study at the same time. Some join study after observance have begun.

• Censoring is an issue to consider…
**Right censoring**: know time of event is greater than some value, but do not know exact value.
• Let’s consider, ‘LEAVE IOWA’ = DEATH

• Now, we’re studying time until a doctor leaves Iowa as a survival analysis.

• We have right-censored data because, at the time of our analysis, not all doctors have left (some have, some haven’t).

• Kaplan-Meier estimator for survivor function...
Kaplan-Meier Estimator

• Estimates a survivor function, $S(t)$, without covariates. $S(t)$ is the probability that time to death is greater than time $t$ (decreasing function).

• Like an empirical CDF: 

\[ \hat{S}(t) = \prod_{i} \left( \frac{n_i - d_i}{n_i} \right)^{t(i)} \leq t \]

- $t_{(i)}$ are the rank order survival times
- $n_i$ number at risk at time $t_{(i)}$
- $d_i$ number of deaths between $t_{(i)}$ and $t_{(i+1)}$
About 51% leave by 15 years out.