Epidemiologic Research in the Danish Registries*

Ryan Seals, ScD
National ID Conference 2015

*Please note that all findings are preliminary
Outline

• Background on our research

• Information on the Danish registries we are using

• Detailed example
Very brief background

- Our group studies risk factors for amyotrophic lateral sclerosis (ALS)
- Rare, fatal neurodegenerative disease
  - Median survival 3 years
- 2-3 per 100,000 individuals
- ~6,000 cases in US every year
- Some risk factors of interest:
  - Medical (injuries), occupational, chemical, residential, pharmaceutical
ALS Etiology

• Very few established risk factors for ALS
  – Age
  – Genetic factors
    • Explain ~70% of familial ALS, but only ~10% of sporadic ALS

• Suspected risk factors
  – Sex (M>F)
  – Smoking
  – Physical trauma
  – Select occupations
  – Industrial chemicals
  – Testosterone
ALS Research in US Data

• Because it is **rare** and **rapidly fatal**, sufficiently powered studies are difficult

• Cancer Prevention Study-II
  – Cohort of 1.1 million individuals
  – 1,156 cases

• National Longitudinal Mortality Study
  – Representative national cohort based on census and current population survey data, followed for cause of death
  – 471 cases
ALS Research in US Data

- Pooled cohort research for nutritional risk factors
  - Harvard, American cancer society, Multiethnic, NIH-AARP cohorts
  - 1 million individuals
  - 995 cases
Danish ID Number

- 10-digit personal identification number (CPR) in use since 1968

- Digits:
  - 1-2 date of birth
  - 3-4 month of birth
  - 5-7 year of birth
  - 8, 9 random numbers
  - 10 gender (male=odd, female=even)

- The establishment of the CPR in 1968 was principally based on two factors:
  1. The growing need for information about common personal data, especially persons residing address, and;
  2. The need for a general personal identification, which could be used everywhere.
Civil Registration System

- Maintains CPR numbers and key demographic data:
  - Name
  - Address
  - Place of birth
  - Vital status (including emigration/disappearance)
  - Date of death
  - Marital status
  - Job title
  - Immediate-family CPR numbers (parents, spouse, children)

- All historic data is retained
Pension Fund

- Established in 1964

- All wage earners and salaried employees contribute (along with employers) to a compulsory pension that begins at age 67

- Contains to-the-day information on when an individual is paid by a particular company

- Classified into Danish version of International Standard Classification of Occupation
  - For epidemiologic consideration, this is company-level code, not occupational roles within companies
National Patient Register

• All public inpatient hospital diagnoses since 1977 & outpatient hospital diagnoses since 1994
  – From 2003 mandatory reporting from private hospitals and clinics (~1% of hospital beds)
• Initially used to monitor hospital activity
  – Since 2000 used for payment

• Data
  – CPR number
  – Hospital Department
  – Date of arrival, Date of departure
  – Action diagnosis
  – Other hospital use variables
Access to data

• Statistics Denmark maintains a research database that links demographic, medical, death, occupational, SES, educational and others for easy research purposes

• Access is available to pre-approved research environments in Denmark
  – Ex. Danish Cancer Society
Example

Occupational formaldehyde exposure and ALS

• We undertook a case-control study of ALS
• Formaldehyde is a neurotoxin linked in prior (problematic) studies to ALS

Study Population:
• All cases of ALS diagnosed from 1982-2009
  – By first ALS diagnosis in National Patient Register
• Each patient matched to 100 healthy controls by age, sex and year of birth
  – From the Central Person Registry
• Epidemiologically, this corresponds to prospectively following the entire Danish population from 1982-2009 for the incidence of ALS
Data sources

- Pension Fund
  - Occupational history
- Patient Register
  - Medical history
- CPR Office
  - Demographic info
- Job Exposure Matrix
  - Formaldehyde exposure
- Information De-identified

ANALYSIS
From occupational history to quantitative chemical exposure

• Job exposure matrix
  – NOCCA-DANJEM
  – Incorporates direct measurements and occupational epidemiologist expert knowledge
  – Levels determined by the product of the probability of exposure and the mean intensity of exposure among the exposed
### The data

<table>
<thead>
<tr>
<th>ID</th>
<th>treat1_preb</th>
<th>nr</th>
<th>status</th>
<th>statdat</th>
<th>lobnr</th>
<th>sex</th>
<th>fsaar</th>
<th>diagdat</th>
<th>SamplingFraction</th>
<th>year</th>
<th>X_j</th>
<th>W_j</th>
<th>X_j_1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.000000000</td>
<td>1</td>
<td>90</td>
<td>05MAR1983</td>
<td>0</td>
<td>1</td>
<td>1987</td>
<td>26JAN1983</td>
<td>1.000000000</td>
<td>1983</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>2</td>
<td>0.000000000</td>
<td>1</td>
<td>90</td>
<td>30NOV1992</td>
<td>70</td>
<td>1</td>
<td>1987</td>
<td>26JAN1983</td>
<td>0.071580119</td>
<td>1983</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>3</td>
<td>0.000000000</td>
<td>1</td>
<td>90</td>
<td>05JUN1999</td>
<td>79</td>
<td>1</td>
<td>1987</td>
<td>26JAN1983</td>
<td>0.071580119</td>
<td>1983</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>4</td>
<td>0.000000000</td>
<td>1</td>
<td>90</td>
<td>09JUL1994</td>
<td>82</td>
<td>1</td>
<td>1987</td>
<td>26JAN1983</td>
<td>0.071580119</td>
<td>1983</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>5</td>
<td>0.000000000</td>
<td>1</td>
<td>90</td>
<td>19SEP2002</td>
<td>87</td>
<td>1</td>
<td>1987</td>
<td>26JAN1983</td>
<td>0.071580119</td>
<td>1983</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>6</td>
<td>0.3174537988</td>
<td>2</td>
<td>90</td>
<td>01JUL1995</td>
<td>0</td>
<td>0</td>
<td>1929</td>
<td>22FEB1993</td>
<td>1.000000000</td>
<td>1993</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>7</td>
<td>0.9122518823</td>
<td>2</td>
<td>90</td>
<td>31DEC2009</td>
<td>16</td>
<td>0</td>
<td>1929</td>
<td>22FEB1993</td>
<td>0.232210820</td>
<td>1993</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>8</td>
<td>0.000000000</td>
<td>2</td>
<td>90</td>
<td>29FEB2000</td>
<td>35</td>
<td>0</td>
<td>1929</td>
<td>22FEB1993</td>
<td>0.115012980</td>
<td>1993</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>9</td>
<td>0.000000000</td>
<td>2</td>
<td>90</td>
<td>27NOV2006</td>
<td>75</td>
<td>0</td>
<td>1929</td>
<td>22FEB1993</td>
<td>0.179054118</td>
<td>1993</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>10</td>
<td>0.000000000</td>
<td>2</td>
<td>90</td>
<td>28APR2007</td>
<td>89</td>
<td>0</td>
<td>1929</td>
<td>22FEB1993</td>
<td>0.185381254</td>
<td>1993</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>11</td>
<td>0.000000000</td>
<td>3</td>
<td>90</td>
<td>13JAN1997</td>
<td>0</td>
<td>1</td>
<td>1929</td>
<td>08JAN1996</td>
<td>1.000000000</td>
<td>1996</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>12</td>
<td>0.000000000</td>
<td>3</td>
<td>1</td>
<td>31DEC2009</td>
<td>4</td>
<td>1</td>
<td>1929</td>
<td>08JAN1996</td>
<td>0.186377725</td>
<td>1996</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>13</td>
<td>0.000000000</td>
<td>3</td>
<td>90</td>
<td>06MAY2004</td>
<td>41</td>
<td>1</td>
<td>1929</td>
<td>08JAN1996</td>
<td>0.158075357</td>
<td>1996</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>14</td>
<td>0.000000000</td>
<td>3</td>
<td>1</td>
<td>31DEC2009</td>
<td>48</td>
<td>1</td>
<td>1929</td>
<td>08JAN1996</td>
<td>0.186377725</td>
<td>1996</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>15</td>
<td>0.000000000</td>
<td>3</td>
<td>90</td>
<td>06JAN2002</td>
<td>72</td>
<td>1</td>
<td>1929</td>
<td>08JAN1996</td>
<td>0.153633889</td>
<td>1996</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>16</td>
<td>0.000000000</td>
<td>4</td>
<td>1</td>
<td>31DEC2009</td>
<td>0</td>
<td>0</td>
<td>1931</td>
<td>16DEC1998</td>
<td>1.000000000</td>
<td>1998</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>17</td>
<td>0.000000000</td>
<td>4</td>
<td>1</td>
<td>31DEC2009</td>
<td>26</td>
<td>0</td>
<td>1931</td>
<td>16DEC1998</td>
<td>0.280581487</td>
<td>1998</td>
<td>0.000000000</td>
<td>0</td>
<td>0.000000000</td>
</tr>
<tr>
<td>18</td>
<td>0.000000000</td>
<td>4</td>
<td>1</td>
<td>31DEC2009</td>
<td>37</td>
<td>0</td>
<td>1931</td>
<td>16DEC1998</td>
<td>0.280581487</td>
<td>1998</td>
<td>0.000000000</td>
<td>1</td>
<td>0.000000000</td>
</tr>
</tbody>
</table>
Results

- 3,650 cases of ALS from 1982-2009
# Results

<table>
<thead>
<tr>
<th>Formaldehyde Exposure **</th>
<th>Controls</th>
<th>Cases</th>
<th>OR* (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Exposure</td>
<td>12869 (88)</td>
<td>3148 (86)</td>
<td>ref</td>
</tr>
<tr>
<td>Any Exposure</td>
<td>1731 (11.9)</td>
<td>502 (13.8)</td>
<td>1.2 (1.1-1.4)</td>
</tr>
<tr>
<td>1&lt;sup&gt;st&lt;/sup&gt; Quartile</td>
<td>425 (2.9)</td>
<td>115 (3.2)</td>
<td>1.2 (0.9-1.4)</td>
</tr>
<tr>
<td>2&lt;sup&gt;nd&lt;/sup&gt; Quartile</td>
<td>447 (3.1)</td>
<td>137 (3.8)</td>
<td>1.3 (1.1-1.6)</td>
</tr>
<tr>
<td>3&lt;sup&gt;rd&lt;/sup&gt; Quartile</td>
<td>406 (2.8)</td>
<td>123 (3.4)</td>
<td>1.3 (1.1-1.6)</td>
</tr>
<tr>
<td>4&lt;sup&gt;th&lt;/sup&gt; Quartile</td>
<td>453 (3.1)</td>
<td>127 (3.5)</td>
<td>1.2 (1.0-1.4)</td>
</tr>
</tbody>
</table>

*All models adjusted for matching factors (age, sex and calendar date), residence, marital status, and SES.

**Quartiles of exposure determined from cases, with cutoffs: 1.48x10^-3 ppm, 4.59x10^-3 ppm, and 1.26x10^-2 ppm.
Other projects

• History of head trauma
• Pharmacologic data
• Residential information
  – Air pollution models
  – Distance from major roadways
  – Distance from industrial sites
• Familial linkages
  – The maternal CPR can be used to link siblings
• Twin studies
  – Ongoing twin registries in most Scandinavian countries
The benefits of registry data

• The registries allow for **full** follow-up of a **well-defined** cohort (Denmark)
• The data are **prospectively** collected
  – Usually for non-research purposes
• The data are **objectively** collected
  – Not subject to individual recall errors
• **Variety** of data sources

• The best future work will combine these benefits of linked registry data with more traditional epidemiologic methods
  – Interviews, biosamples, etc…
Acknowledgments

• The Rose Traveling Fellowship (Thanks Deborah!)

• In Denmark
  – Johnni Hansen, PhD (Danish Cancer Society)
  – Ole Gredal, MD (Danish Neurological Institute)

• Harvard Research Group
  – Marc Weisskopf, PhD, ScD, (PI)
  – Marianne Kioumourtzoglou, ScD

Funding

NIEHS Grant 5R01ES019188-02
Taplin Fellowship
Environmental Epidemiology Training Grant