

# Epidemiologic Research in the Danish Registries\*

---

Ryan Seals, ScD

National ID Conference 2015

\*Please note that all  
findings are preliminary



**HARVARD**  
**T.H. CHAN**  
**SCHOOL OF PUBLIC HEALTH**

# Outline

---

- Background on our research
- Information on the Danish registries we are using
- Detailed example

# Very brief background

---

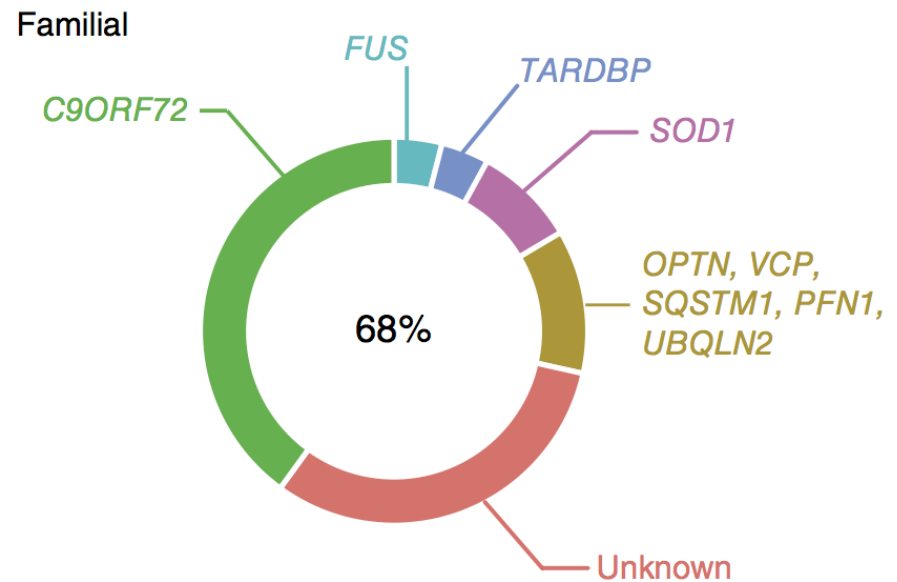
- Our group studies risk factors for amyotrophic lateral sclerosis (ALS)
- Rare, fatal neurodegenerative disease
  - Median survival 3 years
- 2-3 per 100,000 individuals
- ~6,000 cases in US every year
- Some risk factors of interest:
  - Medical (injuries), occupational, chemical, residential, pharmaceutical



# 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

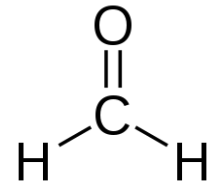
## Occupational formaldehyde exposure and ALS

---

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

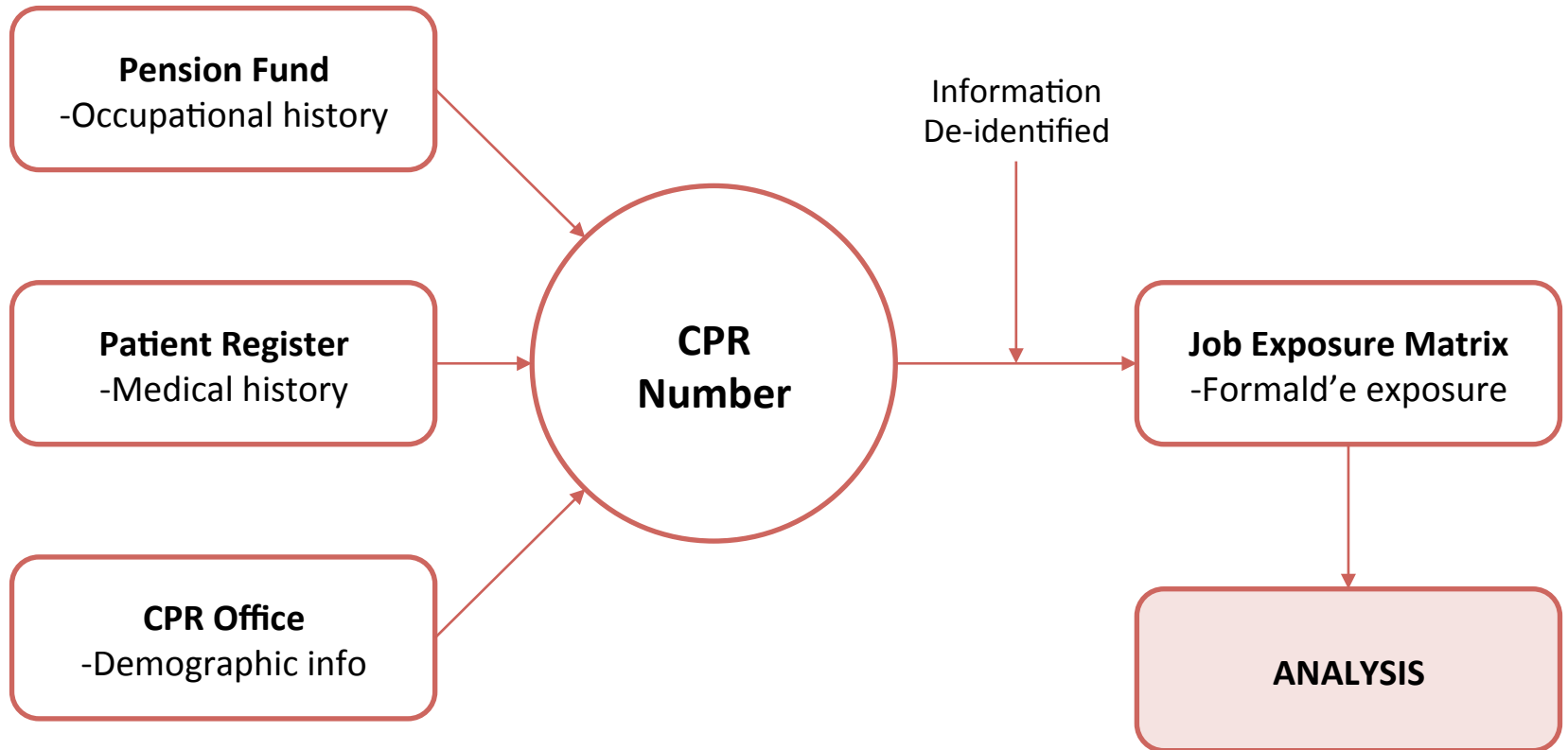
### Study Population:

- 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

---



# From occupational history to quantitative chemical exposure

---

- Job exposure matrix
  - NOCCA-DANJEM
  - Three time periods: 1960-1974, 1975-1984, 1985-
  - 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

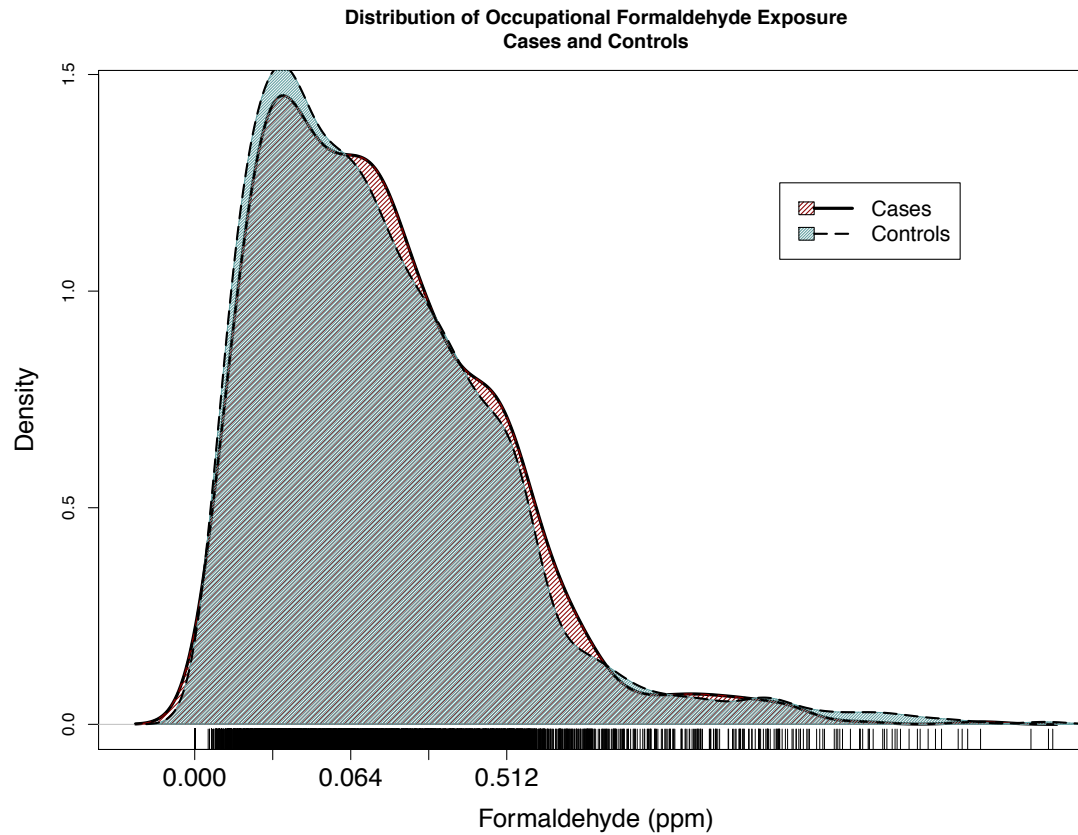
---

	ID	treat1_preb	nr	status	statdat	lobnr	sex	fjaar	diagdat	SamplingFraction	year	X_j	W_j	X_j_1
1	1	0.000000000	1	90	05MAR1983	0	1	1907	26JAN1983	1.000000000	1983	0.000000000	0	0.000000000
2	71	0.000000000	1	90	30NOV1992	70	1	1907	26JAN1983	0.071580119	1983	0.000000000	0	0.000000000
3	80	0.000000000	1	90	05JUN1999	79	1	1907	26JAN1983	0.071580119	1983	0.000000000	0	0.000000000
4	83	0.000000000	1	90	09JUL1994	82	1	1907	26JAN1983	0.071580119	1983	0.000000000	0	0.000000000
5	88	0.000000000	1	90	19SEP2002	87	1	1907	26JAN1983	0.071580119	1983	0.000000000	0	0.000000000
6	102	0.3174537988	2	90	01JUL1995	0	0	1929	22FEB1993	1.000000000	1993	0.000000000	0	0.000000000
7	118	0.9122518823	2	1	31DEC2009	16	0	1929	22FEB1993	0.232210020	1993	0.000000000	0	0.000000000
8	137	0.000000000	2	90	29FEB2000	35	0	1929	22FEB1993	0.115012980	1993	0.000000000	0	0.000000000
9	177	0.000000000	2	90	27NOV2006	75	0	1929	22FEB1993	0.179054118	1993	0.000000000	0	0.000000000
10	191	0.000000000	2	90	28APR2007	89	0	1929	22FEB1993	0.185381254	1993	0.000000000	0	0.000000000
11	203	0.000000000	3	90	13JAN1997	0	1	1929	08JAN1996	1.000000000	1996	0.000000000	0	0.000000000
12	207	0.000000000	3	1	31DEC2009	4	1	1929	08JAN1996	0.186377725	1996	0.000000000	0	0.000000000
13	244	0.000000000	3	90	06MAY2004	41	1	1929	08JAN1996	0.158075357	1996	0.000000000	0	0.000000000
14	251	0.000000000	3	1	31DEC2009	48	1	1929	08JAN1996	0.186377725	1996	0.000000000	0	0.000000000
15	275	0.000000000	3	90	06JAN2002	72	1	1929	08JAN1996	0.153633889	1996	0.000000000	0	0.000000000
16	304	0.000000000	4	1	31DEC2009	0	0	1931	16DEC1998	1.000000000	1998	0.000000000	0	0.000000000
17	330	0.000000000	4	1	31DEC2009	26	0	1931	16DEC1998	0.280581487	1998	0.000000000	0	0.000000000
18	341	0.000000000	4	1	31DEC2009	37	0	1931	16DEC1998	0.280581487	1998	0.000000000	1	0.000000000

# Results

---

- 3,650 cases of ALS from 1982-2009





# Results

---

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

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

\*\*Quartiles of exposure determined from cases, with cutoffs:  $1.48 \times 10^{-3}$  ppm,  $4.59 \times 10^{-3}$  ppm, and  $1.26 \times 10^{-2}$  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)
  - Marianthi Kioumourtzoglou, ScD

## Funding

**NIEHS Grant 5R01ES019188-02**

**Taplin Fellowship**

**Environmental Epidemiology Training Grant**