DATA CORE MEETING

Observational studies in dementia and the MELODEM initiative

V. Shane Pankratz, Ph.D.

Outline

Theme: Overcoming Challenges in Longitudinal Research

- **Challenges in observational studies**
- **MELODEM**
- **Discussion:**

  Are there collaborative opportunities we could take advantage of as data cores?

Challenges in observational studies

- Existential question: Why do we do what we do?

- To even begin answer, we need to address two other questions:
  1. What do we do?
  2. What are the goals to be addressed by what we do?

Challenges in observational studies

What do we do?

Data Cores are instrumental in compiling data for NACC repository

- **The study population: The UDS** (http://www.alz.washington.edu/WEB/study-pop.html)
- Subjects with dementia, MCI, and intact cognition – reflecting total enrollment at ADCs since 2005.
- Data are collected by trained clinicians from subjects as and self-designated informants
- Diagnoses are assigned in per protocol at each ADC (process can vary by center)
- NACC subjects are not from a population-based sampling (referral or volunteer participants)
  - Many ADCs require agreement to autopsy as a precondition for enrollment

Challenges in observational studies

What do we do?

- **Data collection for the UDS** (http://www.alz.washington.edu/WEB/study-pop.html)
- UDS data collection began in September 2005.
- UDS data comprise a standardized evaluation of subjects enrolled in ADCs
- UDS protocol requires annual, longitudinal follow-up as long as the subject is able to participate (late-stage subjects may be followed for autopsy-only after discontinuing UDS visits)
- Focus is Alzheimer's disease, but Centers also enroll and collect data on other
  - e.g. vascular dementia, dementia with Lewy bodies, and frontotemporal lobar degeneration
- UDS includes more than 700 variables
  - demographics, behavioral status, cognitive testing, medical history, family history, clinical impressions, and diagnoses.
- Information is collected from in-person office visits and telephone calls, as well milestone forms documenting subject death and drop-out

Challenges in observational studies

What do we do?

- **Data description** (http://www.alz.washington.edu/WEB/data-descript.html)

	Minimum data set (MDS)	Uniform Data Set (UDS) (LONGITUDINAL)	Neuropathology Data Set (NP)
Years covered	1984 - 2005	Sept. 2005 - present	1984 - present
Study subjects	Enrollees followed at ADCs (with or without dementia)	Enrollees followed at ADCs (with or without dementia)	Subjects who died and underwent autopsy
Approx. # of subjects*	74,397	29,004	13,429
Approx. # of variables	67	725	85
Method of data collection	Mainly abstracted retroactively from ADC medical records	Collected prospectively by clinicians, neuropsychologists, and other ADC research personnel, using up to 18 standardized forms at each visit. Some forms also have Spanish-language and telephone versions.	Standardized neuropathology form, completed by neuropathologist
Time period covered for each subject	Mainly status on last ADC visit; some variables also capture initial-visit status	Initial visit and each annual follow-up visit, plus milestones such as death or dropout	Status of brain at autopsy
Topics covered (brief list)	Demographics, cognitive status, clinical dementia diagnosis, selected clinical manifestations, comorbid conditions, MMSE score, vital status, primary neuropathological diagnosis (if died and had brain autopsy)	Sociodemographics on subject and informant, family history, dementia history, neurological exam findings, functional status, neuropsych-ological test results, clinical diagnosis, whether imaging testing done, ApoE genotype	Demographics, date of death, primary and secondary neuropathological diagnoses, presence/absence of neuropathological features of most major dementias, APOE genotype, brain weights

Challenges in observational studies

Why do we do it?

- **A harder question . . .**

1. Describe natural history?
2. Understand natural history?
   - Implications differ for rare vs. common diseases vs. unaffected individuals
3. Identify correlates of disease onset/worsening?
4. Prove causal links between risk factors and disease?
5. Understand disease etiology?
6. Other reasons?

Challenges in observational studies

How can we appropriately answer research questions?

- **Asking the right question is key:**
  - Far better an approximate answer to the *right* question, which is often vague, than an *exact* answer to the wrong question, which can always be made precise. John Tukey. *The future of data analysis*. Annals of Mathematical Statistics 33 (1), (1962), page 13.
- **Even when we ask the right question, we need to work to conduct an appropriate study:**
  - The combination of some data and an aching desire for an answer does not ensure that a reasonable answer can be extracted from a given body of data. John Tukey. *Sunset salvo*. The American Statistician 40(1).
  - Led Ronald Fisher to the idea of using randomization as a way to design studies to test causal hypotheses
  - Mosteller liked to say: “You can *only* prove causality with statistics.”
- **When we have the right study design, we can easily define the right analysis**
  - What if the study design is complicated, or the question difficult?
  - This is often the case when we are conducting longitudinal observational research

Challenges in observational studies

- **Issues common to all epidemiological research**
  - Unmeasured confounding
  - Selective participation
  - Missing data
  - etc.
- **Some issues may be more prevalent or severe in dementia research**
  - Measurement error (especially lack of a “gold standard” outcome measure)
  - End of life effects
  - etc.

Challenges in observational studies

- One challenge: different analytical approaches are often chosen to address the same question with the same/similar data
- Many studies do not use the analytic methods most likely to produce unbiased and precise effect estimates
  - Different views about the most appropriate approaches
  - Distinct disciplinary traditions in epidemiology, biostatistics, neuropsychology, and neurology
  - Different statistical methods available to researchers due to software and other barriers

Challenges in observational studies

- Lack of a shared logical basis for selecting particular methods has at least two major adverse consequences
  - It becomes difficult to quantitatively summarize results across different studies
  - It becomes difficult to understand whether different associations are attributable to differences in study design or in statistical methods
- Even if different approaches are ultimately selected (e.g. due to untestable background assumptions) it can be valuable to specify a set of standardized sensitivity analysis approaches

MELODEM

MEthods for LOngitudinal research in DEMentia

- Initiative formed in 2012
- Initial intent was to study the difficulties mentioned above, with the ultimate goal of achieving greater unity in the selection of analytical methods in the conduct of research on dementia risk and cognitive aging
- Did not focus on issues in study design (although perhaps implicit in selection of analytical methods)

Leadership Group:

Geneviève Chêne, Carole Dufouil, M. Maria Glymour

Supported by

Fondation Plan Alzheimer

MELODEM

- Current Goals
  1. Achieving better consensus on the extent to which challenges might affect results (bias, bias/precision tradeoffs, other issues?)
  2. Selecting, when possible, the best methodologic approaches to derive unbiased and precise effect estimates
  3. Identifying barriers to uptake of preferred methods
  4. Identifying areas of high priority for research methods development

MELODEM

- Focus Areas
  1. Handling selection due to study participation, attrition, and mortality
  2. Dealing with measurement of exposure and outcomes
  3. Addressing time-varying exposures and time-varying confounding
  4. Specifying longitudinal models
  5. Analyzing high-dimensional data (e.g. neuroimaging, genetic)

MELODEM

"Selection"

- Selective study enrollment and differential attrition can lead to apparent associations that are non-causal, or mask associations that are
  - *Differential attrition and survival after enrollment*
  - In studies of dementia, poor cognition has a recognized influence on post-enrollment attrition
  - *Differential study enrollment*
  - In studies of dementia, likelihood of enrollment can be influenced by the exposure of interest and cognition (or their determinants)
  - *Solution(s)?*

MELODEM

"Measurement"

- Do our measures of cognitive function truly reflect disease state?
  - *Practice of retest effects*
  - Repeated exposure to the same cognitive testing regimen tends to result in improved performance on the tests
  - *Reliability/measurement error*
  - Testing context, interviewers, transient fluctuations in participants, may influence performance on cognitive tests
  - *Unequal interval scaling/Ceiling and floor effects*
  - e.g. Does a decline from 25 to 24 on the MMSE mean the same as a decline from 20 to 19?
  - *Solution(s)?*

MELODEM

"Time scale / age"

- What is the best time scale, and what does one do with time-varying characteristics?
  - *Multiple measurements and measurement error*
  - Issues arise when the precision with which measurements are made over time varies
  - *Time-varying exposures or confounders*
  - Changes in risk factors over time may influence cognitive outcomes
  - Risk for "reverse causation"
  - *Unequal interval scaling/Ceiling and floor effects*
  - Does a decline from 25 to 24 on the MMSE mean the same as a decline from 20 to 19?
  - *Solution(s)?*

MELODEM

"Longitudinal model specification"

- What are approaches which might be used to study questions in cognition?
  - *Convergence between within-person change and between-person age difference*
  - Is it appropriate to assume a common age trajectory for all?
  - *Terminal decline and non-linear age trends*
  - What is the most appropriate form of trends in cognition as a function of age? Do they change in the time just prior to death?
  - *Other ideas?*

MELODEM

"High dimensional data"

- Availability of data-rich sources opens new challenges and opportunities.

- *Multiple testing*
- Control of false discovery is critical. What is the best approach, especially when measurements might be correlated?
- *More measurements than subjects and over-fitting*
- Use dimension reductions techniques prior to analysis?
- Use analytical approaches which are designed to reduce over-fitting issues (e.g. lasso)?

- The challenge is to extract the meaningful information from these data and use it to understand dementia etiology.

MELODEM

- Thoughts on goals?
  1. Achieving better consensus on the extent to which challenges might affect results (bias, bias/precision tradeoffs, other issues?)
  2. Selecting, when possible, the best methodologic approaches to derive unbiased and precise effect estimates
  3. Identifying barriers to uptake of preferred methods
  4. Identifying areas of high priority for research methods development

MELODEM

- Thoughts on Focus Areas?
  1. Handling selection due to study participation, attrition, and mortality
  2. Dealing with measurement of exposure and outcomes
  3. Addressing time-varying exposures and time-varying confounding
  4. Specifying longitudinal models
  5. Analyzing high-dimensional data (e.g. neuroimaging, genetic)

MELODEM

or Data Cores? or Both?

- Should we contribute?
- What can we focus on?
  - Any particular strengths we could bring to bear from within ADC program?