Fostering open access for Alzheimer’s and dementia research: Challenges and the NACC experience

Walter A. Kukull, PhD
Professor of Epidemiology
Director, National Alzheimer’s Coordinating Center
University of Washington, Seattle WA, USA
Types of database sharing (results)

• Journal “supplemental data”
  • “Here are the specific data files from which our results came...check our work if you like!”

• GWAS Results
  • “We gathered DNA from 20,000 cases and controls, enrolled in similar but independent studies and conducted a GWAS, using an Illumina ###; here are the genotype results for your use...” (dbGap, NIAGADS et al)
  • Curated PLINK files etc.
Data sharing (2)

• Unique measurements, e.g. ADNI
  • Highly standardized imaging data
  • Clinical trial-like enrollment protocol

• Rare populations, e.g. DIAN
  • Autosomal dominant AD families

• Large population-based cohort studies
  • Work with PI to obtain data
  • Replicate findings or analyze new questions
Issues for the researcher

• Use Individual studies
  • Internally standardized and consistent
  • Is Power adequate, i.e., number of subjects

• Combine, i.e. pool, data from several studies
  • Are collected data really the same across studies
  • How to “harmonize”

• Standardized, large, multi-site, extensive, data collection and f/u
  • Define comparison groups; ask novel questions you choose
Challenges for the researcher seeking data

• Differing study designs may affect potential results?

• Are data highly structured and standardized?
  • Within and/or Between studies
  • Are measures of data reliability/validity available?
  • AD biomarker assays are notoriously variable
Implications:
Study design and data collection

• Data combined or analyses done without regard to limitations of study design or design differences may provide spurious findings

• In addition to treatment/exposure and outcome, the factors associated with them must be accurately gathered to adequately assess confounding;
  • For example, other risk/prognostic factors
Implications:
Study design and data collection

• Systematic flaws in subject enrollment or data measurement that lead to study group differences ARE a “permanent” bias;

• Unreliable/misclassified data mask effects;
  • Suppose 30% of “controls have AD pathology?

• Non-Valid data provide non-valid results despite large “N” or any level of statistical significance obtained.
Challenges for “effective” data interpretation or combination of studies?

• Do variables mean different things between clinicians or studies, but appear to be similar on CRF or DED?
  • Clinical judgment v. ambiguity?
  • Facts not in evidence?

• Is extensive documentation included?
  • Only DED or are example CRFs or study protocol available?

• Are Meta-Data included?
  • Is Study Principal Investigator(s) available for consultation?
Implications of study data for “effective” analysis by researchers

- Non-standardized data should not be combined casually
- Detailed documentation, meta-data and expert consultation positively impact validity and interpretation of analytic results.
  - Applies to individual study data as well as combination data files
Challenges of Data combination and “Data mining” for AD/dementia studies

• Individual study flaws and data inaccuracies will still lead to spurious results despite statistical significance;

• Clinical data are gathered with “error”

• Beware Un-hypothesized, multiple comparisons
  • Even “p < 1.0 \times 10^{-13}”, does not overcome design and data collection flaws

• GWAS : a special case?
  • Common outcome measurement by “machine”
Implications of Data combination and “Data mining” for AD/dementia studies

• Without attention to documentation et al, “study combination” results must be carefully scrutinized to be trusted;
• “Pure” data mining results, without understanding of content or acquisition of data, should be viewed with considerable skepticism;
Who should have access to data I will share?

- Research group members only?
- Any person?
  - Should “registration” be required (primarily as use information and output tracking)?
  - Anonymous use permitted?
  - Third party distribution and editing permitted?
- Any “bona fide” researcher?
  - Professional involved in scientific research.
- Only the people I like?
What form should the data take? Where would it be located?

• All study data in a series of downloadable files, along with DED, on request and approval?
• Only data determined by study Exec Committee?
• Public site, all study files, unmonitored?
• Private website; password protected?
• Public Study website, specific data by request?
  • Encrypted data analysis files distributed
• Are all data De-identified or anonymized?
Accessing NACC data

• Data are available to anyone
• All CRF, “guidebooks”, DED are available to view or download; direct data query facility
• Expert consultation with NACC analyst to tailor analysis data file to researcher’s request
  • NACC constructs file and sends it to investigator usually within 2 WD; updates and expansion on request; all free!
• We do not “dump” entire raw database to researchers
  • Large, longitudinal, relational database and multiple form changes with time: we do the hard work for you.
Welcome to NACC’s researcher website

REQUEST NACC DATA HERE
- Perform an online query
- Request a custom data table
- Request a data file for analysis
- Locate tissue

Click on one of the links in the box to request data now. No password or account is required. No affiliation with the NIA Alzheimer’s Disease Centers Program is necessary.

For more information about researcher responsibilities and processing your request, as well as other NACC data information, please access the links in the far left column.

Submit your manuscript or abstract before publication
Researchers who use NACC data in their manuscripts or abstracts must submit their work to NACC before publication. This is required to ensure that NACC-related manuscripts are tracked and that the NACC grant (U01 AG016976) is credited. Please download and review the Checklist for Authors here and submit your manuscript or abstract here.

In order to comply with the NIH Public Access Policy, all authors using NACC data must ensure the proper submission of their published work to PubMed Central (PMC). For more information on this process, please review the PDF “Resources for navigating PMC submission.”

To check the status of a manuscript, and as a reminder to the author to follow up with PMC if necessary, NACC will send authors an email reminder six months after the manuscript is submitted to NACC, and every three months after that, until we have received the PMCID.
The NACC Database

Uniform Data Set (UDS)

Minimum Data Set (MDS)

Neuropathology Data Set (NP)

Centers Data

Collaborative Projects Data Sets

Other AD Grants Archived Data Sets

Biospecimens & Imaging (BIDSS)

<table>
<thead>
<tr>
<th>IMAGING</th>
<th>BIOSPECIMENS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research sMRI</td>
<td>CSF</td>
</tr>
<tr>
<td>FDG - PET</td>
<td>Genetics/APOE</td>
</tr>
<tr>
<td>Amyloid PET</td>
<td>Plasma/Serum</td>
</tr>
<tr>
<td>CT Scans</td>
<td>DNA</td>
</tr>
<tr>
<td></td>
<td>Brain Tissue</td>
</tr>
</tbody>
</table>
The UDS — clinical data

UDS临床数据

- Demographics: subject and informant 人口统计资料：受试人及陪伴者
- Family history/pedigree data 家族史/世系资料
- Current medications 目前用药
- Health history 健康史
- Physical: height, weight, BP, HR, reported vision/hearing 体征：身高，体重，血压，心脏率，视力/听力
- Hachinski scale and CVD
- UPDRS (motor section)
- CDR
- NPIQ
- Geriatric Depression Scale 老年忧郁程度
The UDS — clinical data
UDS临床数据

- Functional activities (FAQ) 功能性活动
- Physical/neuro overall findings 体征/神经整体发现
- Clinician judged symptoms/onset 临床判断症状/发病
- Neuropsychological battery 神经心理系列
  - MMSE
  - Logical memory (Immed /delay) 逻辑记忆
  - Digit span (forward/back) 数字间距（前进/后退）
  - Category fluency (animal/veg) 类别流畅程度（动物/植物）
  - Trails A&B 轨迹
  - WAIS-R Digit Symbol WAIS-R数字符号
  - Boston Naming test (30 item) Boston 命名测试
- Clinician diagnoses 临床诊断
## Cognitive status

<table>
<thead>
<tr>
<th>Clinical category</th>
<th>UDS (any)</th>
<th>Deaths</th>
<th>Neuropath</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not impaired</td>
<td>9,565</td>
<td>561</td>
<td>254</td>
</tr>
<tr>
<td>MCI or other impaired</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amnestic</td>
<td>3,816</td>
<td>380</td>
<td>155</td>
</tr>
<tr>
<td>Non-amnestic</td>
<td>1,009</td>
<td>105</td>
<td>52</td>
</tr>
<tr>
<td>Not MCI</td>
<td>1,201</td>
<td>86</td>
<td>33</td>
</tr>
<tr>
<td>Demented</td>
<td>12,185</td>
<td>3,446</td>
<td>1,912</td>
</tr>
<tr>
<td>Total</td>
<td>27,776</td>
<td>4,578</td>
<td>2,406</td>
</tr>
</tbody>
</table>

Numbers as of the March 1, 2013 data archive
Primary clinical diagnosis for dementia

<table>
<thead>
<tr>
<th>Clinical diagnosis</th>
<th>UDS (any)</th>
<th>Deaths</th>
<th>NP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer’s disease*</td>
<td>9,223</td>
<td>2,355</td>
<td>1,260</td>
</tr>
<tr>
<td>Vascular dementia</td>
<td>296</td>
<td>91</td>
<td>48</td>
</tr>
<tr>
<td>Dementia with Lewy bodies</td>
<td>633</td>
<td>301</td>
<td>168</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>178</td>
<td>71</td>
<td>36</td>
</tr>
<tr>
<td>FTLD**</td>
<td>1,365</td>
<td>461</td>
<td>300</td>
</tr>
<tr>
<td>Other dementia</td>
<td>490</td>
<td>167</td>
<td>100</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12,185</strong></td>
<td><strong>3,446</strong></td>
<td><strong>1,912</strong></td>
</tr>
</tbody>
</table>

Numbers as of the March 1, 2013 data archive

*includes probable and possible Alzheimer’s
** FTLD includes bvFTD, SD, PPA, CBD, PSP, and Picks
Thank You to the OECD and the conference organizers for inviting me.
Thanks also to all my colleagues at the ADCs and at NACC and to the patients and families who participated in this research;
And of course thanks to NIA for funding NACC!

NACC is funded by the National Institute on Aging (U01 AG016976) and located in the Department of Epidemiology at the University of Washington School of Public Health, Walter A. Kukull, Ph.D., Director.
Copyright© 2013 University of Washington.