Penn Integrated NeuroDegenerative Disease Database (INDD)

INQuery - Search and You Will Find

Rui Tong, Data Manager
University of Pennsylvania Alzheimer’s Disease Core Center
Integrated Neurodegenerative Disease Database (INDD)

Clinical Databases
- AD
- ALS
- PD
- FTLD

Supporting Databases
- Genetics
- Biofluid
- Neuro path
Building an integrated neurodegenerative disease database at an academic health center

Sharon X. Xie\textsuperscript{a,b,c,d,e,g}, Young Baek\textsuperscript{b,d,e,f}, Murray Grossman\textsuperscript{b,d,e}, Steven E. Arnold\textsuperscript{c,d,e,f}, Jason Karlawish\textsuperscript{b,d,e,h}, Andrew Siderowf\textsuperscript{d,e,f}, Howard Hurtig\textsuperscript{d,e,f,j,k}, Lauren Elman\textsuperscript{d,e,f,l}, Leo McCluskey\textsuperscript{d,e,f,l}, Vivian Van Deerlin\textsuperscript{b,d,e,j,m}, Virginia M.-Y. Lee\textsuperscript{b,d,e,m}, John Q. Trojanowski\textsuperscript{b,d,e,j,k,m}

\textsuperscript{a}Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{b}Alzheimer’s Disease Core Center, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{c}Center for Frontotemporal Degeneration, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{d}Institute on Aging, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{e}Center for Neurodegenerative Disease Research, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{f}Department of Neurology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{g}Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{h}Penn Memory Center, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{i}Division of Geriatrics, Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{j}Morris K. Udall Parkinson’s Disease Research Center of Excellence, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{k}Parkinson’s Disease and Movement Disorder Clinic, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{l}Amyotrophic Lateral Sclerosis Center, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
\textsuperscript{m}Department of Pathology and Laboratory Medicine, University of Pennsylvania School of Medicine, Philadelphia, PA, USA

INDD Database

- Relational Database
- Microsoft SQL Server 2008 R2 with ASP.NET MVC 3
- Contains \geq 300 Tables
- Data from \geq 17000 Research Subjects
- Millions of Data Records
INQuery - Search and you will find

Web-based database query tool, enabling end-users to query and extract data without the need of database administrators

Objectives
• Web-based: Universal access
• Easy to use with intuitive interface
• Includes advanced features to filter, merge, sort data

Methods
• Using stages to perform queries
• Filtering, sorting between stages
• Merge tables based on specific date ranges
• Sort by earliest or latest visit / test dates
## INQuery - Terminology

<table>
<thead>
<tr>
<th><strong>INDD</strong></th>
<th>Integrated NeuroDegenerative Disease (Database)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INQuery</strong></td>
<td>INDD Querying System</td>
</tr>
<tr>
<td><strong>INDDID</strong></td>
<td>Unique identifier for individual patients in the INDD Database System</td>
</tr>
<tr>
<td><strong>Dataset</strong></td>
<td>2 dimensional table layed out in columns and rows (active / working)</td>
</tr>
<tr>
<td><strong>Table</strong></td>
<td>Collection of records stored in the database (database)</td>
</tr>
<tr>
<td><strong>Record</strong></td>
<td>A single row within a dataset, table</td>
</tr>
<tr>
<td><strong>Field</strong></td>
<td>Name of columns represented in a table</td>
</tr>
<tr>
<td><strong>Data types</strong></td>
<td>Int, float, decimal, string (varchar), bit (boolean), date</td>
</tr>
<tr>
<td><strong>Primary key</strong></td>
<td>A field or fields in a table that can uniquely identify all records in a table</td>
</tr>
<tr>
<td><strong>Filter</strong></td>
<td>Ability to remove records based on specified criteria</td>
</tr>
<tr>
<td><strong>Logical [AND]</strong></td>
<td>Combination of two or more filters operating together. Both must be true in order to result in true statement</td>
</tr>
<tr>
<td><strong>Logical [OR]</strong></td>
<td>Combination of two or more filters operating independently of each other. Either one of them can be true to result in true statement</td>
</tr>
<tr>
<td><strong>Merge / Join</strong></td>
<td>Process of combining dataset with data from a table</td>
</tr>
<tr>
<td><strong>Left outer join</strong></td>
<td>Process of leaving the current dataset post-merge regardless of a matching record is present in the joining table</td>
</tr>
</tbody>
</table>
INQuery - Layout

1. Reset, Commit, Export, Search
2. Table List, Fields
3. Paste initial dataset below (First row must be labels)
4. Total Patients: 0, Total Rows: 0
INQuery – Generating Queries

1. Start out with a question you want to ask the database.
2. Import the initial dataset if you have one.
3. Add a table from the table list to the working dataset.
4. Use field filters / field checkboxes to narrow down to desired dataset.
5. Add additional tables and merge them to the current active dataset.
6. Further refine the active dataset using filters / checkboxes.
7. Repeat Steps 5, 6.
INQuery – Live Demo

• Male Patients from ADC
• Diagnosis of AD or MCI
• MMSE of $\geq 20$ at their latest visit
• CSF
• APOE Status
INQuery – Data Security

Data Security

• Enforced by levels of access rights
• Different users have access to different sets of tables
• We can choose which table or field to be published to which group.
• Logical tables are published using Views to separate data by their sensitivity

Views

• Views are virtual tables.
• Result set of a stored query on data.
• Can represent a subset of data.
• Can join and simplify multiple tables.
• Can act as aggregated tables.
• Do not take up actual storage space.
INQuery Activities

- Pre-INQuery: ~3 queries per day

- Post-INQuery: ~Average of 153 queries per week / 21 queries per day
Some Studies Inspired by INQuery

**Abnormal serine phosphorylation of insulin receptor substrate 1 is associated with tau pathology in Alzheimer’s disease and tauopathies**

Mark Yarchan, Jon B. Toledo, Edward B. Lee, Zoe Arvanitakis, Hala Kazi, Li Ying Han, Natalia Lonjeva, Virginia M.-Y. Lee, Sangwon F. Kim, John Q. Trojanowski, Steven E. Arnold

**Cerebrovascular atherosclerosis correlates with Alzheimer pathology in neurodegenerative dementias**

Mark Yarchan, Sharon X. Xie, Mitchel A. Kling, Jon B. Toledo, David A. Wolk, Edward B. Lee, Vivianna Van Deerlin, Virginia M.-Y. Lee, John Q. Trojanowski and Steven E. Arnold

**A platform for discovery: The University of Pennsylvania Integrated Neurodegenerative Disease Biobank**


**Comparative Survey of the Topographical Distribution of Signature Molecular Lesions in Major Neurodegenerative Diseases**

Steven E. Arnold, Jon B. Toledo, Dina H. Appleby, Sharon X. Xie, Li San Wang, Young Baek, David A. Wolk, Edward B. Lee, Bruce L. Miller, Virginia M.-Y. Lee, and John Q. Trojanowski
Acknowledgements

• Young Baek (Former Data Manager)
• John Trojanowski
• Steve Arnold
• Sharon Xie
• All Colleagues of Penn ADCC
• NIA Funding (AG 101024)
• NACC, an inspiration of current work
• Patients and Their Families
Thank You!

Q & A