

NeuroIE:  
Extracting Neuroimaging Study  
Results from the Literature

**Yong Gao, Ph. D.**

Mass General Hospital, Harvard Medical School

**David N. Kennedy, Ph. D.**

University of Massachusetts Medical Center

# Motivation

- **Neuroimaging Techniques (MRI, fMRI and PET)**
  - brain images are acquired and analyzed to study the structure and function of the brain
- **Neuroimaging Literature**
  - Report correlation between the brain structure and behavior observations or disease diagnosis for groups of subjects demographically characterized
- **Databasing the Neuroimaging Literature**
  - Internet Brain Volume Database (IBVD)
  - SumsDB
- **Manual database curation is the bottleneck**

# Tabular Data Extraction: Example

**Table 1** Descriptive statistics for unadjusted volumes

Brain region	Autistic group		Control group		Between group comparisons	
	Mean brain volume (cm <sup>3</sup> )	SD	Mean brain volume (cm <sup>3</sup> )	SD	ES	<i>P</i> value
Total brain volume	1453.96	136.98	1367.44	106.17	0.670	0.077
Diencephalon	23.34	2.49	21.16	1.18	0.969	0.004
Cerebral white matter	441.04	59.14	383.94	52.24	0.917	0.010
Cerebellum	155.03	12.31	144.88	8.60	0.864	0.011
Globus pallidus–putamen	16.47	1.91	15.06	1.17	0.810	0.028
Brainstem	22.94	2.72	21.30	2.14	0.642	0.057
Caudate	10.92	1.66	10.05	1.20	0.581	0.134
Cerebral cortex	755.30	81.38	739.70	76.62	0.199	0.675
Hippocampus–amygdala	14.75	1.98	14.98	1.63	-0.129	0.697

Overall *F*-test of group differences was obtained using multivariate GLM for correlated data, controlling for age and scanner [ $F(8,31) = 2.52, P = 0.031$ ]. *P* values refer to univariate tests of total brain and regional volume differences. ES = (autistic mean – control mean)/pooled SD.

Table in the article for volumetric measurements for brain regions

grp	vol id	structure	hemisphere	volume mean	volume std	units	flags
<a href="#">1221</a>		Brain	total	1453.96	136.98	cc	
<a href="#">1222</a>		Central Gray Matter	total	23.34	2.49	cc	
<a href="#">1223</a>		Cerebral White Matter	total	441.04	59.14	cc	
<a href="#">1224</a>		Cerebellum	total	155.03	12.31	cc	
<a href="#">1225</a>		Lenticulate	total	16.47	1.91	cc	
<a href="#">1226</a>		Brainstem	total	22.94	2.72	cc	
<a href="#">1227</a>		Caudate	total	10.92	1.66	cc	
<a href="#">1228</a>		Cerebral Cortex	total	755.3	81.38	cc	
<a href="#">1229</a>		Hippo-Amygdala Complex	total	14.75	1.98	cc	

Data for the autistic group manually extracted and reported in IBVD

# Full Text Data Extraction: Example

## Material and methods

### *Subjects*

Quantitative volumetric analysis was performed on brain MRIs of 32 boys (17 autistic, 15 normal control) between 7 and 11 years of age. All 17 autistic boys had performance intelligence quotients (IQs) >80. All qualifying male subjects with MRIs of adequate quality were included in the analysis. Girls were excluded because the imaged sample included only two autistic girls with performance IQ >80. Autistic children were recruited between 1985 and 1988 by clinical referral or by participation in school special needs programs (Rapin, 1996). The control subjects were recruited specifically to the imaging arm of the study and were eligible if they had normal developmental history without seizures or

```
species: Human
in/ex vivo: in vivo
number of subjects: 17
number left handed: not recorded
number right handed: not recorded
number male: 17
number female: 0
diagnosis: Autism
age (mean +- std): 9 +- 3
minimum age: 7
maximum age: 11
```

# Ontology-Driven Information Extraction

- NeuroLex (from NIF)
  - Used to identify brain regions in tables
- Neuroimaging Study Ontology (In development)
  - Provides guidance and constraints for analyzing tables, document structure and natural language sentences
  - Functions as the intermediate framework for representing the extracted knowledge (in RDF)

# NeuroIE System Design

