

Comparing machine-learned and diagnosis-driven groupings of children in the 2011 National Survey of Children's Health

Background

- Developmental disabilities are often complex heterogeneous conditions with multiple concomitant diagnoses
- How similar would ML-formed groups be to existing categories of developmental disability?
- Is grouping children by similarity of diagnoses and condition labels more clinically useful than grouping them by condition labels alone (CP, Autism, Tourettes)

Methods

- 80,000 responses from the 2011 NSCH
- Compare traditional disability labels to 7 machine-learning (ML) groups based on 17 diagnoses/condition labels using K-means clustering
- Create prevalence matrices to display results

Results

- ML classifies every child to only 1 group
- Diagnoses more concentrated in ML groups
- No "CP" in ML groups; diagnoses more important than etiology
- Children in the groups are much more similar after ML groupings, as measured by mean distances

Conclusions

- Children are more clinically similar when grouped on all diagnoses rather than traditional labels
- Some diagnoses may have sub-groups (ADHD) which are different in character
- Some terms may overlap substantially in meaning or be different levels on a common theme (developmental delay, intellectual disability, learning disability)
- Identifying children based on disabilities and diagnoses may improve treatment

	Single-variable groups: prevalence of diagnoses						
	CP	Autism	TBI	Dev Delay	Int Disab	Speech	Tourette's
N	244	1,509	276	2,780	944	3,472	143
behavior	0.11	0.28	0.13	0.23	0.30	0.14	0.28
cp	1.00	0.02	0.31	0.06	0.11	0.04	0.02
adhd	0.20	0.42	0.20	0.37	0.37	0.23	0.57
autism	0.13	1.00	0.13	0.30	0.33	0.22	0.29
anxiety	0.13	0.37	0.23	0.26	0.28	0.16	0.42
asthma	0.17	0.16	0.19	0.19	0.19	0.17	0.18
brain	0.35	0.02	1.00	0.06	0.10	0.03	0.04
depress	0.05	0.12	0.08	0.10	0.14	0.06	0.18
devdelay	0.73	0.56	0.57	1.00	0.86	0.49	0.27
diabetes	0.00	0.01	0.02	0.01	0.01	0.01	0.01
hearing	0.09	0.05	0.10	0.09	0.13	0.10	0.06
intdisab	0.43	0.21	0.35	0.29	1.00	0.19	0.07
joints	0.62	0.11	0.41	0.18	0.27	0.12	0.13
learning	0.66	0.70	0.57	0.77	0.94	0.54	0.45
seizure	0.31	0.06	0.29	0.09	0.17	0.06	0.06
speech	0.59	0.51	0.44	0.61	0.70	1.00	0.24
tourette	0.01	0.03	0.02	0.01	0.01	0.01	1.00

Table 1: Prevalence of condition labels and symptoms for children in condition-defined groups

	Machine learning groups: prevalence of diagnoses						
	1	2	3	4	5	6	7
N	62,895	5,989	1,759	1,440	3,138	3,186	1,758
behavior	0.00	0.01	0.02	0.62	0.11	0.09	0.23
cp	0.00	0.00	0.01	0.00	0.00	0.00	0.08
adhd	0.00	0.00	0.07	0.58	1.00	0.56	0.33
autism	0.00	0.00	0.06	0.13	0.03	0.06	0.43
anxiety	0.01	0.02	0.04	0.89	0.07	0.08	0.26
asthma	0.00	1.00	0.14	0.24	0.13	0.10	0.19
brain	0.00	0.00	0.01	0.01	0.00	0.01	0.08
depress	0.00	0.01	0.01	0.75	0.03	0.03	0.06
devdelay	0.00	0.01	0.15	0.15	0.02	0.10	0.97
diabetes	0.00	0.01	0.00	0.01	0.01	0.01	0.01
hearing	0.01	0.02	0.08	0.05	0.02	0.03	0.11
intdisab	0.00	0.00	0.01	0.04	0.00	0.02	0.45
joints	0.02	0.04	0.04	0.10	0.03	0.04	0.23
learning	0.00	0.04	0.15	0.42	0.00	1.00	0.96
seizure	0.00	0.01	0.02	0.03	0.01	0.02	0.12
speech	0.00	0.00	1.00	0.07	0.00	0.04	0.84
tourette	0.00	0.00	0.00	0.02	0.01	0.01	0.01

Table 2: Prevalence of condition labels and symptoms for children in ML-defined groups

Group	Description
1	No developmental disability/no diagnoses
2	Asthma only
3	Speech impairment only
4	ADHD + Mood problems
5	ADHD only
6	ADHD + learning disabilities
7	Multiple: learning, dev delay, speech

Table 3: Potential descriptive labels for ML-defined groups of children

➤ **Key takeaway: ML-derived groupings may help clinicians develop treatment plans for children with disabilities by discovering subgroups within – and forming new groups across – diagnostic categories.**

