Enhancing Epidemiological Research using Medical Big Data **through Compact Visualization**

Arriel Benis, Moshe Hoshen

1 Clalit Research Institute, Chief Physician's Office, Clalit Health Services

BACKGROUND

Epidemiological studies generally include large to massive quantities of data such as administrative and medical data collected over many years. These data can be considered Big Data.
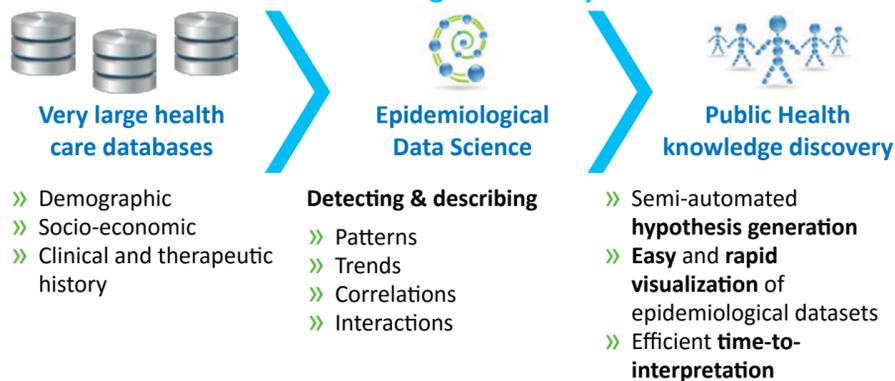
Epidemiological Data Science addresses detection and description of
» Patterns
» Trends
» Correlations
» Interactions

for supporting Public Health Knowledge Discovery by:
» semi-automated hypothesis generation
» concurrent visualization of results
» efficient time-to-interpretation

Epidemiological Data Science can facilitate Public Health Knowledge Discovery

Very large health care databases
» Demographic
» Socio-economic
» Clinical and therapeutic history

Epidemiological Data Science
Detecting & describing
» Patterns
» Trends
» Correlations
» Interactions

Public Health knowledge discovery
» Semi-automated **hypothesis generation**
» **Easy** and **rapid visualization** of epidemiological datasets
» Efficient **time-to-interpretation**

EPIDEMIOLOGICAL DATA SCIENCE OBJECTIVES

Use Compact Visualization to define:
» Relationship between birth season – (youngest, middle, and oldest thirds) and prescription of ADHD medication;
» Relationship changes over time;
» The profiles of children who have similar trends between birth season and year of first purchase of ADHD medication over time.

## METHODS

Compact Visualization is built as follows:
» Graphical representation of the hierarchical clustering:
   1. A distance matrix between pairs of ADHD incidence values for all available discretized attributes describing the children population by birth season
   2. A dendrogram (Unsupervised Learning) based on the distance matrix
» Heatmap:
   1. Columns reflect birth season and year of first purchase
   2. Rows reflect Hierarchical Clustering of the children's socio-demographics
   3. Color gradient reflects this relationship for each birth season by year of study

Experimental data ADHD in young population

	6-17 years old	• Age groups split into two-year bands • Birth Season - youngest, middle and oldest thirds
	2006 - 2011	
	1,013,149 children	CLALIT 100 years
	Socio-Demographics	• Number of siblings (0-1, 2-3, and 4 or more) • Socio-economic status (Low, Medium, High) • Country of Birth • Religion/Ethnicity
	Age at First Prescription	Year of first purchase

RESULTS

**Classical Approach: birth season and first purchase of ADHD medication, 2006-2011**
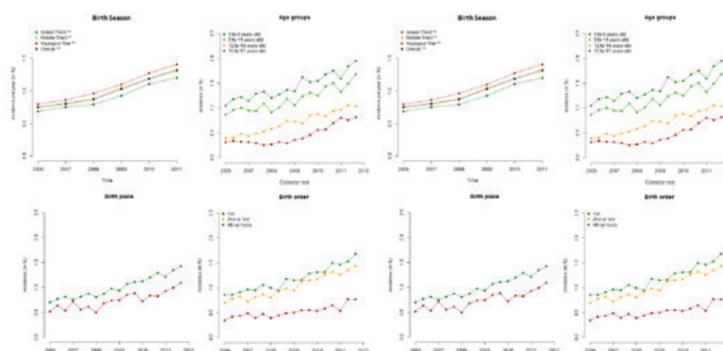
Compact Visualization: birth season and first purchase of ADHD medication, 2006-2011
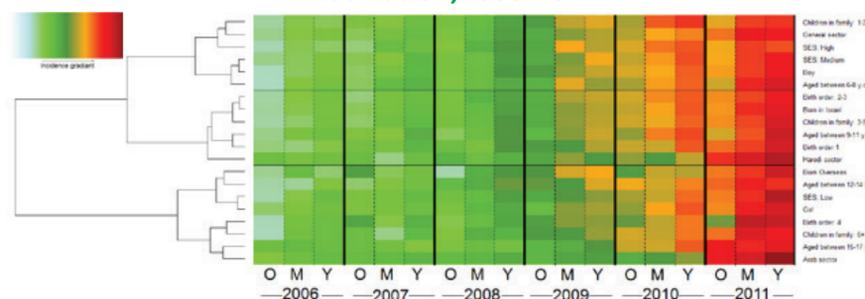
Summary

	Classical Approach	Compact Visualization
Hypothesis	Defined by the **domain expert**	**Semi-automated**
Time-To-Interpretation	Days	Hours
Number of figures	8	1
Easy-To-Understand Easy-To-Read		
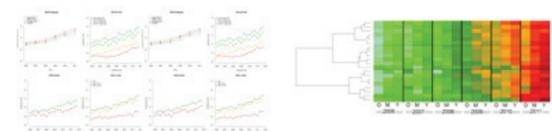
Interpretation

» The youngest children in class purchase medication more often than the oldest children.
» Incidence of medication purchase increases from 2006 to 2011.
» Three main profiles were detected:
   1. Jewish boys aged 6-8 from medium SES small families (1-2 children)
   2. Older siblings (1st to 3rd in birth order) pre-teens of medium-sized families (3-5 children)
   3. Younger siblings (birth order 4 or more) Arab or immigrant girls from low SES large families (6 or more children)

CONCLUSION

Perspectives and Health Policy Implications
Compact Visualization allows for:
» Identification of clusters of attributes with similar trends
» Simultaneous representation of interesting profiles of patients
can also:
» Support public health research reporting and health policy development
» Provide the medical decision makers with an integrated data-analysis tool

References

[1] Hoshen, M. B., Benis, A., Keyes, K. M., and Zoëga, H. (2016) Stimulant use for ADHD and relative age in class among children in Israel. Pharmacoepidemiol Drug Saf, doi: 10.1002/pds.3962

[2] Hoshen MB., Benis A., Zoega H., Keyes KM. (2015) Relative age in class and ADHD in school children – Diagnosis and medication patterns– Intra-annual and inter-annual disparity [Hebrew], The 11th annual Health Policy Conference organized by the Israel National Institute for Health Policy Research, Tel-Aviv, Israel (BEST POSTER AWARD)

Clalit Research Institute
Real people. Real data. Real change.

CLALIT 100 over years
The Best For Your Family