Tracking global outbreak of Hepatitis Delta Virus

By Christopher John Peterson, Himanshi Sharma and Dayanidhi Tandra

THE UNIVERSITY OF UTAH

Project Purpose

Our efforts with the project were targeted on identifying the diseases that supplemented the growth of HDV and help in tracking the outbreak in HDV.

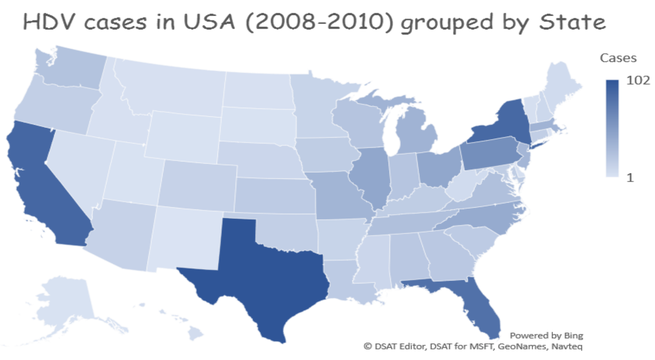
2) Bootstrapping

* Randomly sample data to estimate case distribution
* Resulted in 583 significant diseases amongst HDV patients

3) Logistic Model

Chances of getting infected by HDV

With	Female	Male
No disease	9.81%	11.76%
only HBV	99.99%	99.99%
Only Heart Diseases	18.69%	15.80%
Only Diabetes	17.35%	14.62%



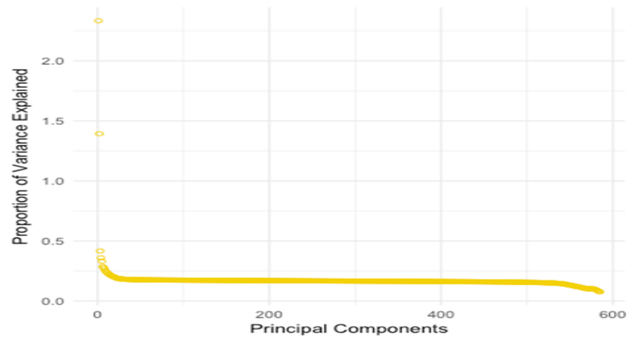
*Costal area are highly affected

Predictive Modelling

Input
Demographic Info of patients along with dummy for 580 significant diseases

Output
The odds of having HDV given demographics and diseases

1) PCA



2) SMOTE for Class Balancing

	No (HDV – cases)	Yes (HDV+ cases)
Original	1359926	2788
After resampling with SMOTE	5576	5576

Conclusion

Bootstrapping yielded too many potential diseases. Logistic regression model helped find the odds of having HDV and can help with reducing the list of disease that supplement the growth of HDV as we continue to work with Prof. Weller.

Exploratory Data Analysis

1) Frequent Items by Misra-Gries

Top 24 Frequent items, Using Misra-Gries Algorithm (k=100)

S.No	Diagnosis Code Pairs	Frequency Counter
1	2724_4019	74
2	7030_70715	6
3	5715_71590	4
4	7030_78720	9
5	7032_25000	104
6	7030_41400	36
7	2809_7030	181
8	311_7032	41
9	3051_7032	35
10	2720_4019	37
11	7032_78720	10

Acknowledgements

Datasets are provided by Weller Lab, Salt Lake City.

For more visualization please scan this.