

Utilizing computational tools to visually explore multivariable datasets: a Lassa fever case study in Nigeria



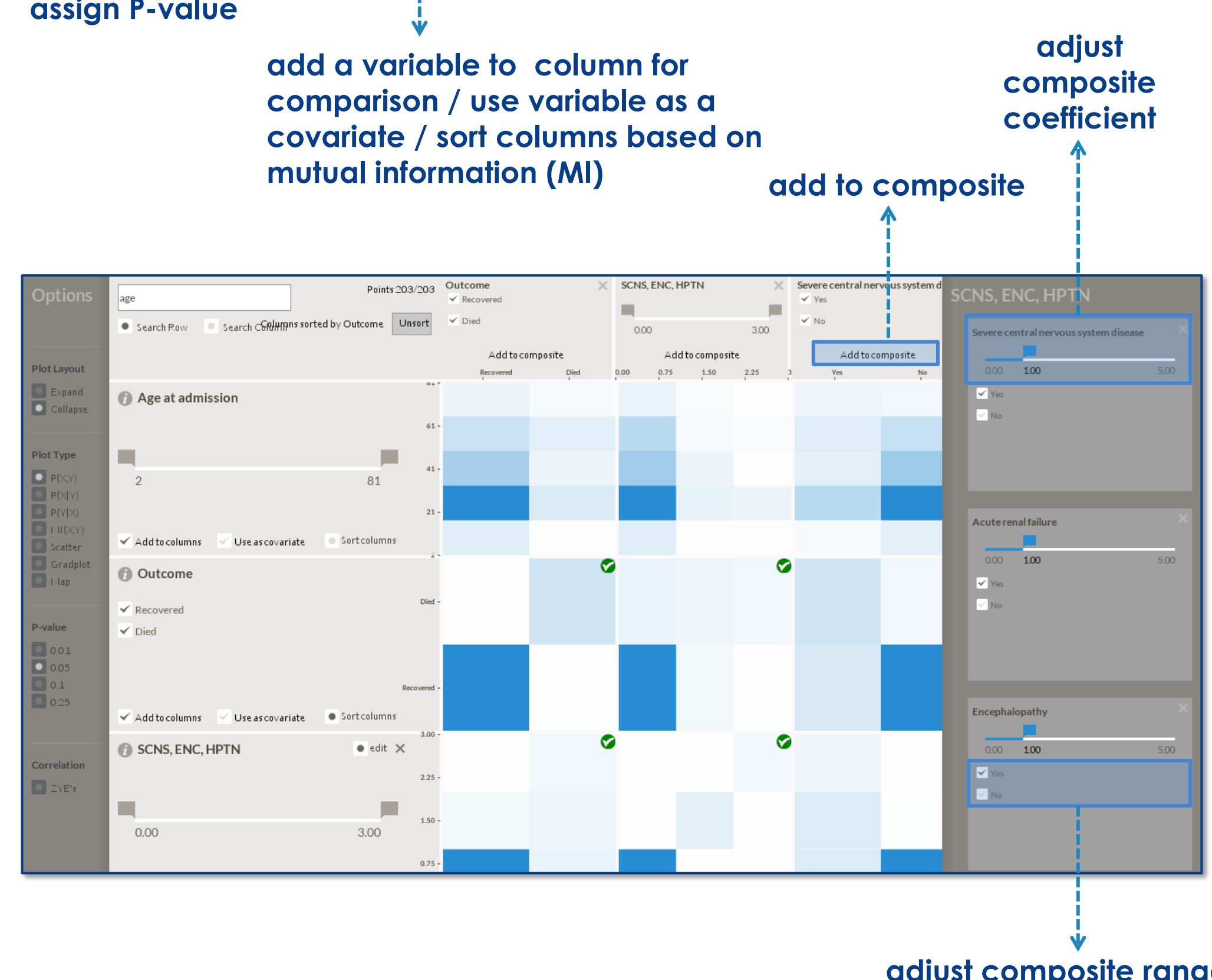
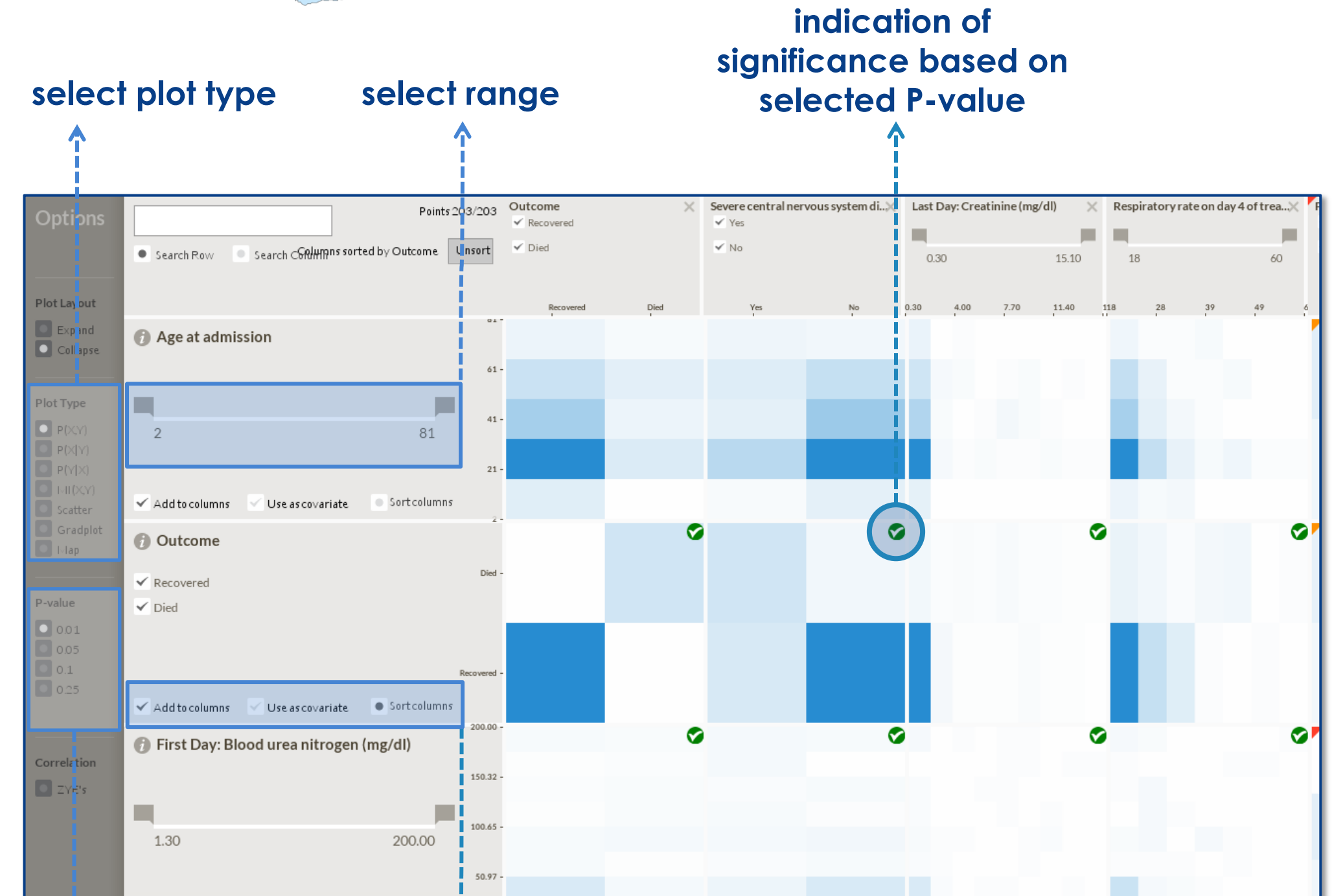
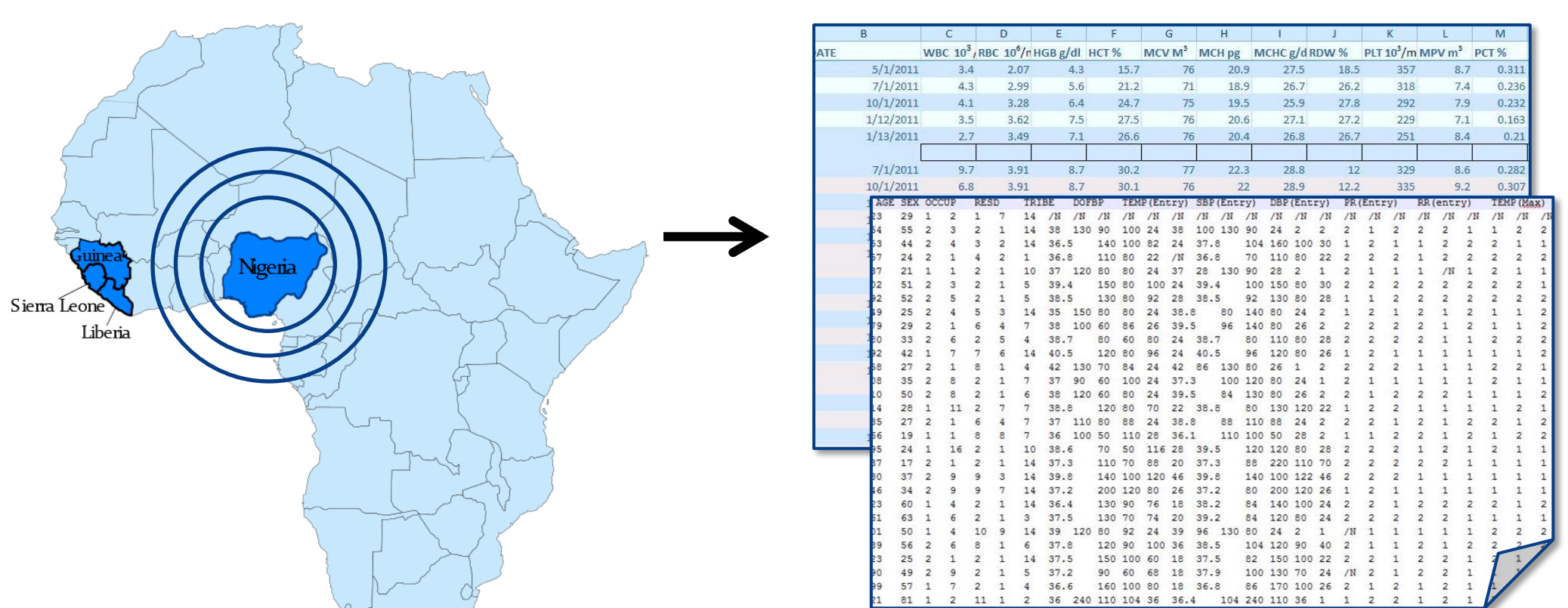
Sara Asad, Andres Colubri, Pardis Sabeti

FAS Center for Systems Biology, Harvard University, Cambridge, MA 02138

Background

Lassa fever is a viral hemorrhagic fever endemic in West African countries including Nigeria, Sierra Leone, and Guinea. The Lassa virus is a negative-strand RNA virus carried in small African rodents, the multimammate rat, which serve as the virus' main reservoir. An estimated 100,000 to 300,000 people are infected by the Lassa virus every year in West Africa. The Irrua Specialist Teaching Hospital (ISTH) in Nigeria has on-site diagnostics and treatment for Lassa fever patients. Clinical data of Lassa patients from January 2011 until June 2013 is utilized in this case study.

Zye Interface



Abstract

The Irrua Specialist Teaching Hospital (ISTH) in Nigeria documents extensive and comprehensive epidemiological data of patients in the Lassa fever ward. Information from this epidemiological data is used in determining the case-fatality rate associated with Lassa fever patients, and most importantly, finding which parameters strongly predict patient outcome. This project presents a visualization software, Zye, which allows users to efficiently scan large datasets through visual analysis. In our results, clinical diagnosis of severe central nervous system disease

(SCNS), encephalopathy (ENC), acute renal failure (ARF), and hypotension (HYPO) most closely prognosticate outcome. Zye provides users an intuitive understanding of associations in multivariable datasets using mutual information ranking and an exploratory visual interface. The ability to extricate meaningful associations and generate novel hypotheses from large datasets has important implications for future epidemiological studies and more broadly, health sciences research.

Results

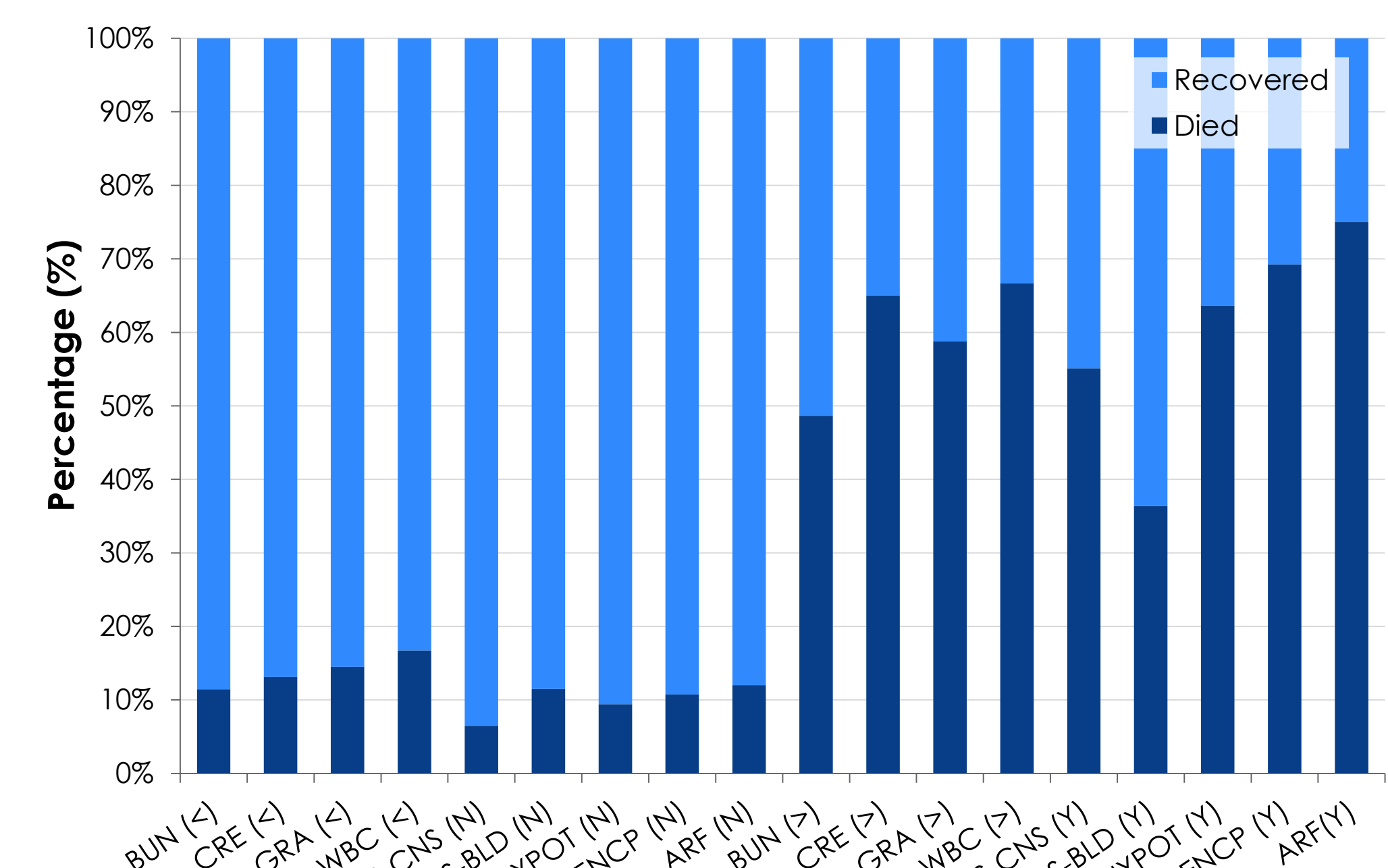
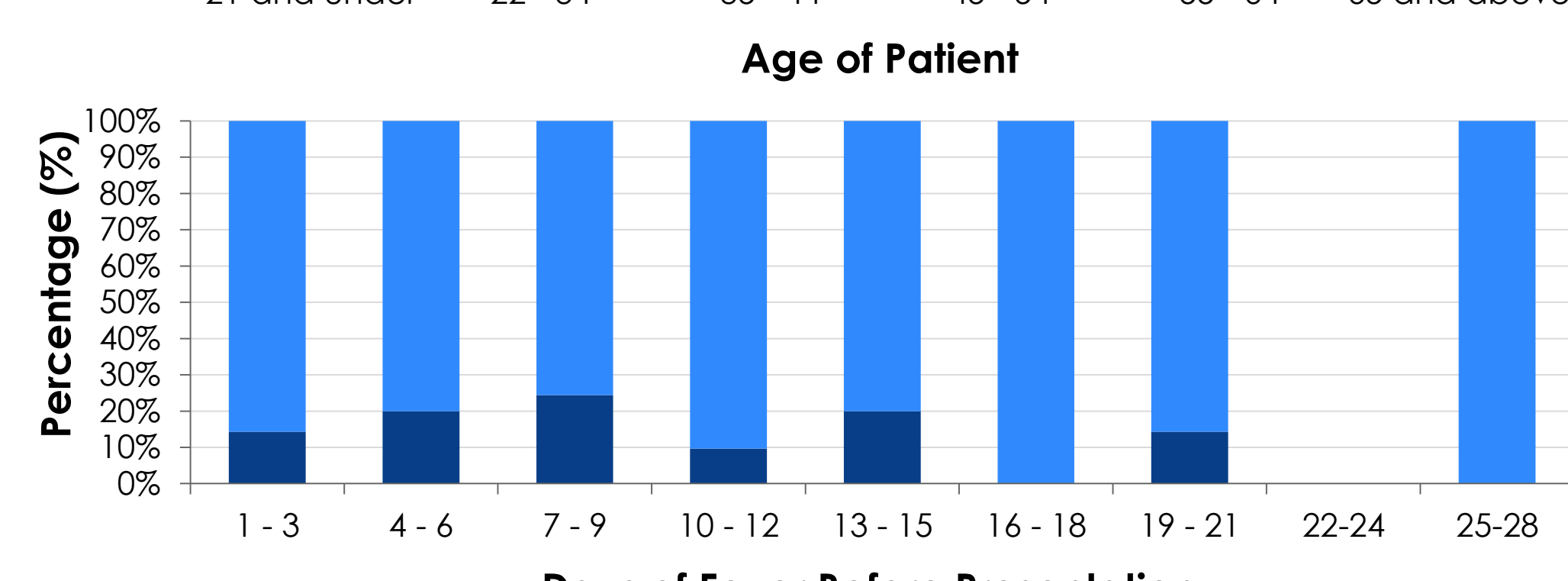
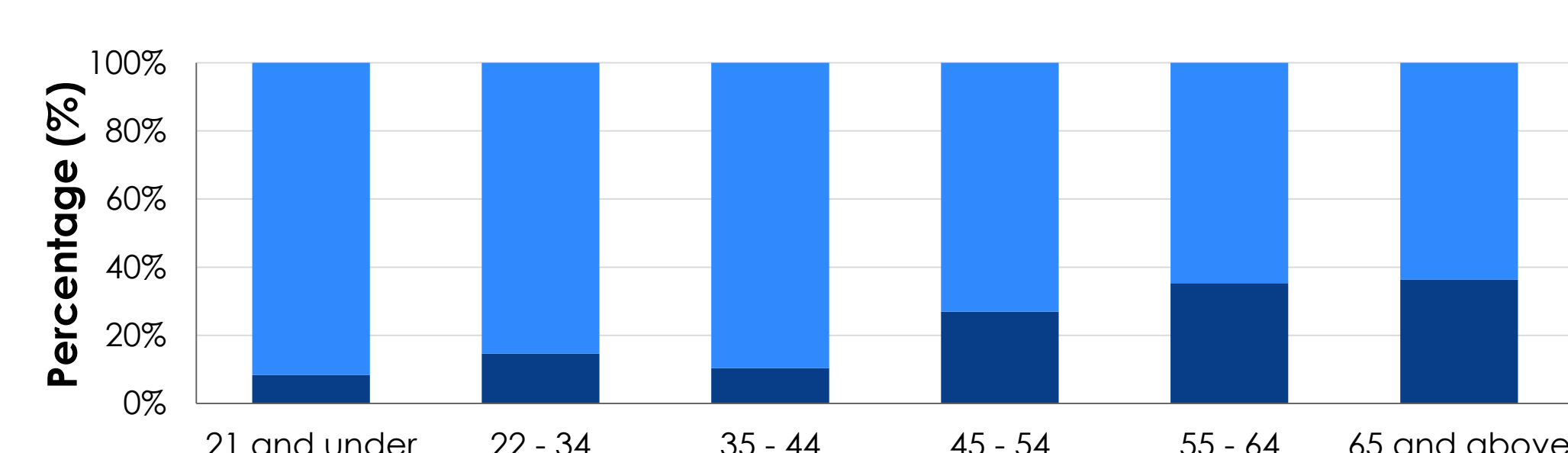


Figure 1. Case-fatality rate for the top nine variables from Zye correlation ranking. The threshold for blood urea nitrogen is above 31.0 (mg/dl), creatine above 6.00 (mg/dl), granulocytes above 9.50 (%) and white blood cell concentration above 18.0 ($10^3/mm^3$).



Figures 2, 3. The effect of a patients' age and the days of fever before presentation at the hospital on CFR.

Predicted Class	Actual Class		Sensitivity and Specificity	
	Recovered	Died	Recovered	Died
Recovered	133 80.1%	6 3.6%	95.7%	4.3%
Died	2 1.2%	25 15.1%	92.6%	7.4%
		98.5%	80.6%	95.2%
		1.5%	19.4%	4.8%

Figure 4. The outcome confusion matrix using SCNS, ARF, ENC, and HYPO followed by its corresponding sensitivity and specificity.

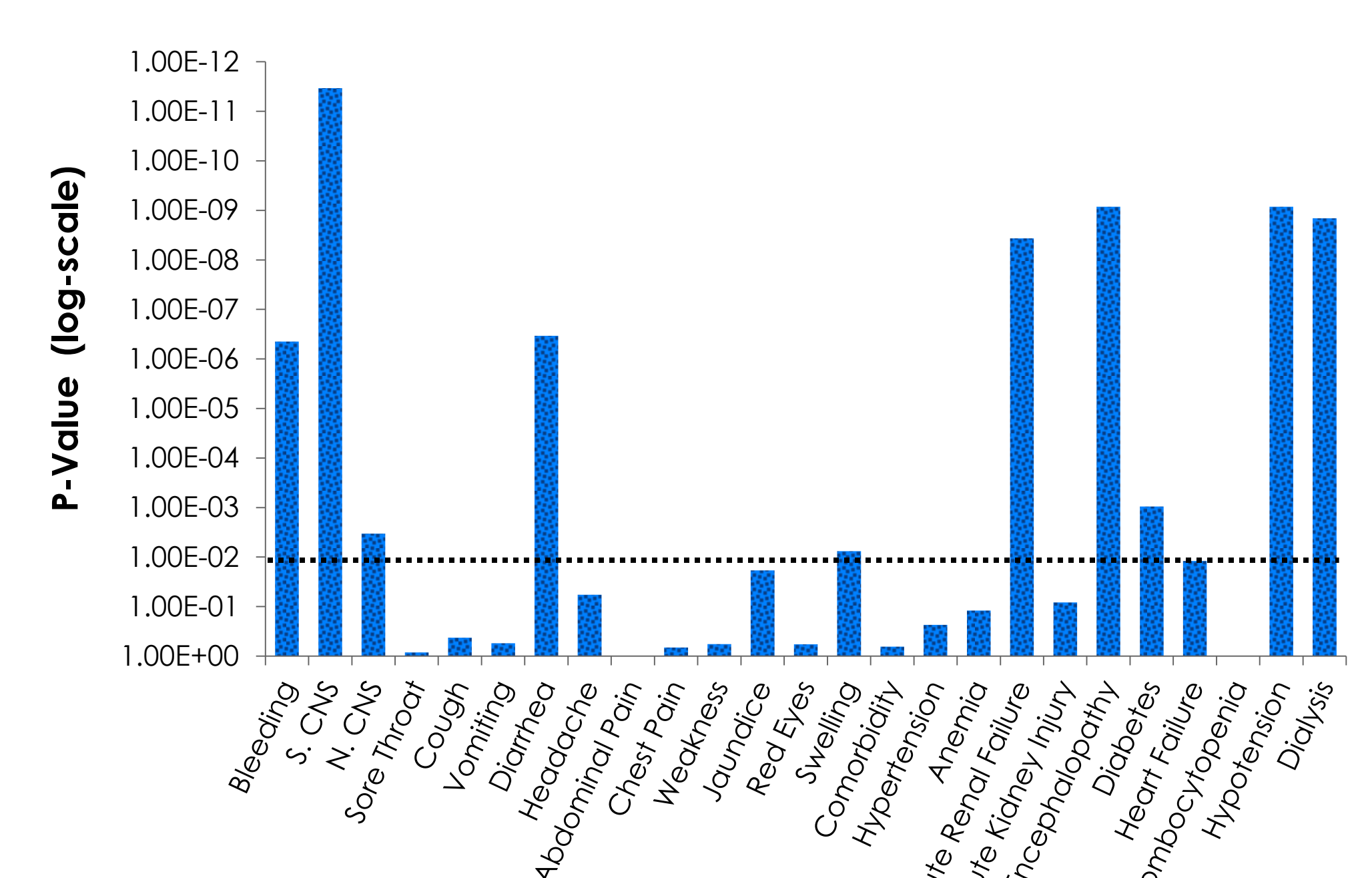


Figure 6. P-values for the degree of significance between discrete clinical variables and patient outcome using Fisher's Exact Test. Variables above the line indicate a P-value less than 0.01.

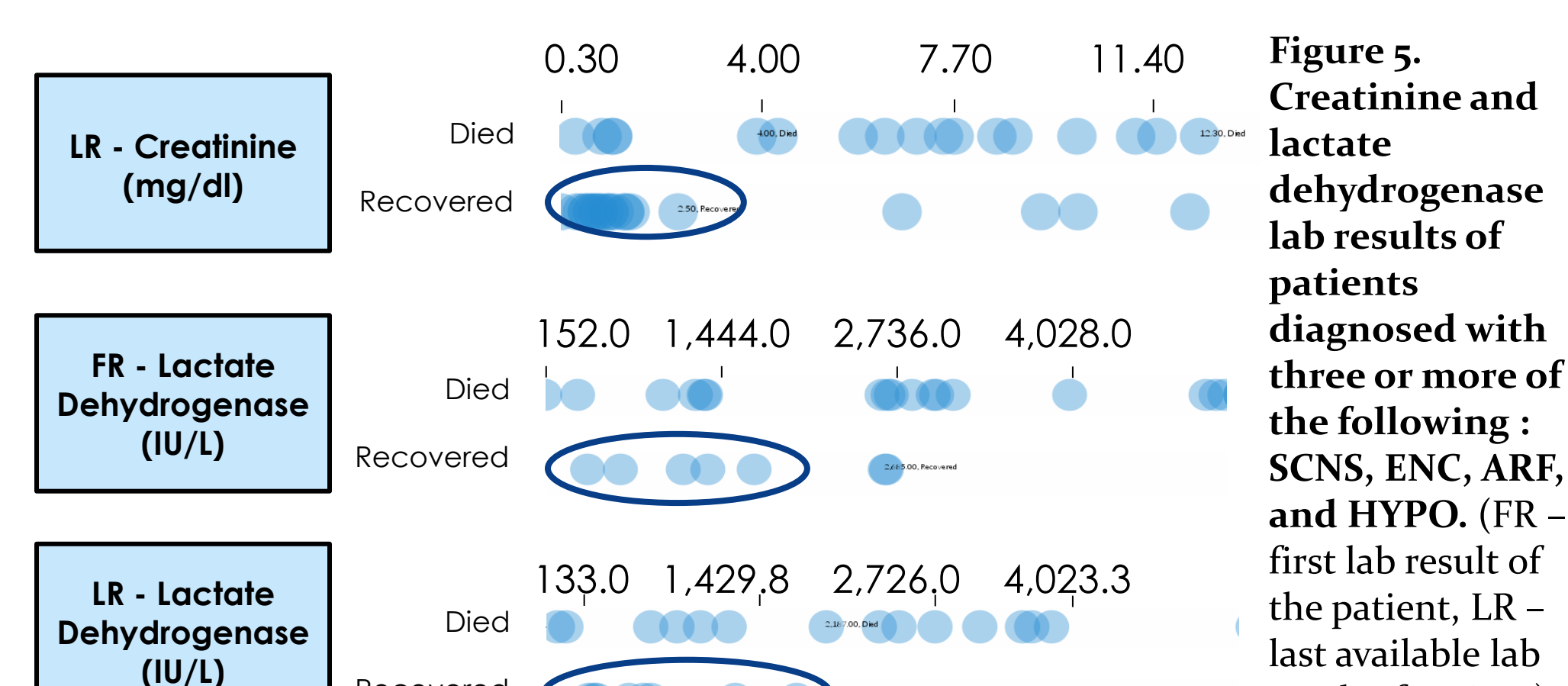


Figure 5. Creatinine and lactate dehydrogenase lab results of patients diagnosed with three or more of the following: SCNS, ENC, ARF, and HYPO. (FR - first lab result of the patient, LR - last available lab result of patient)

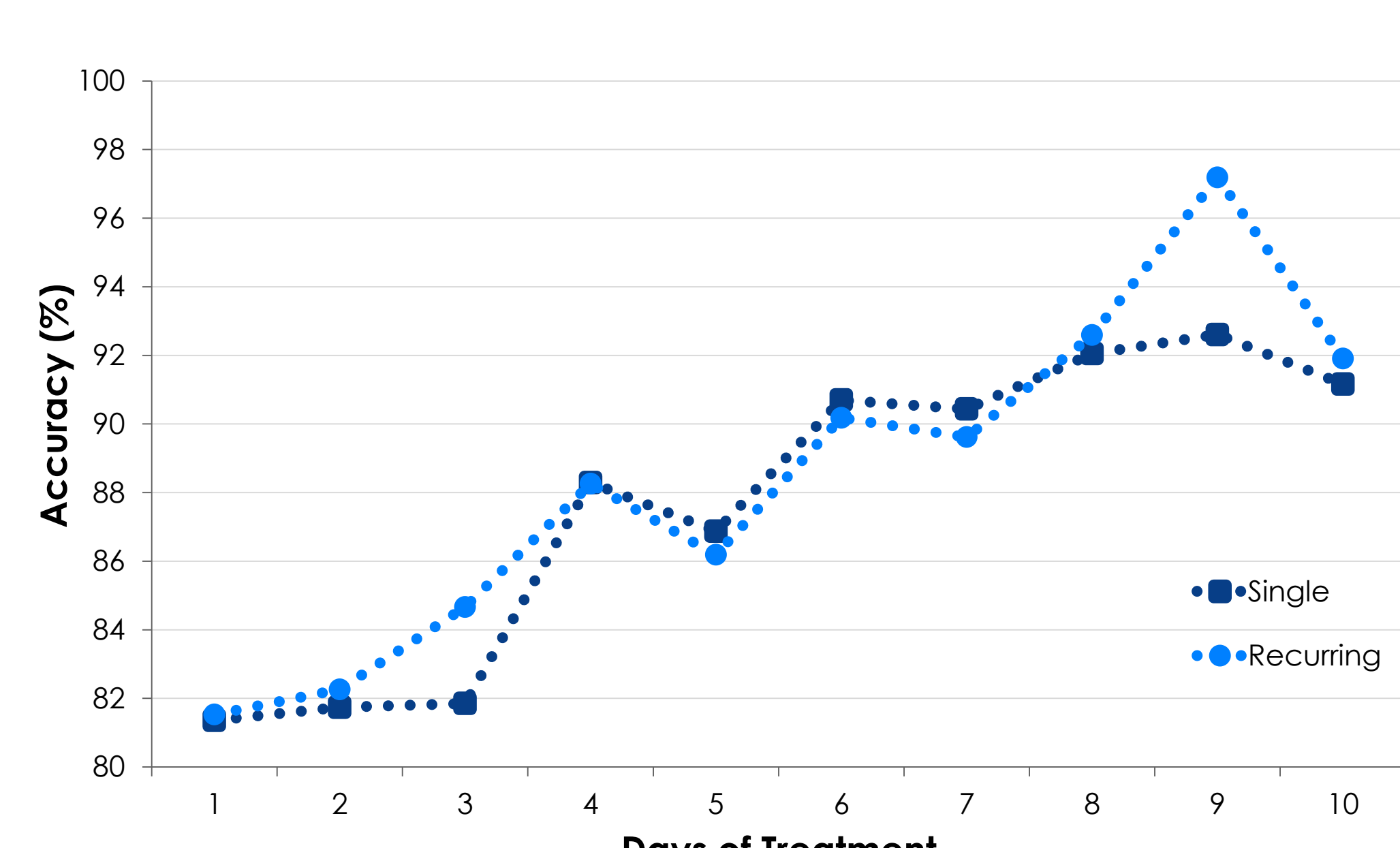
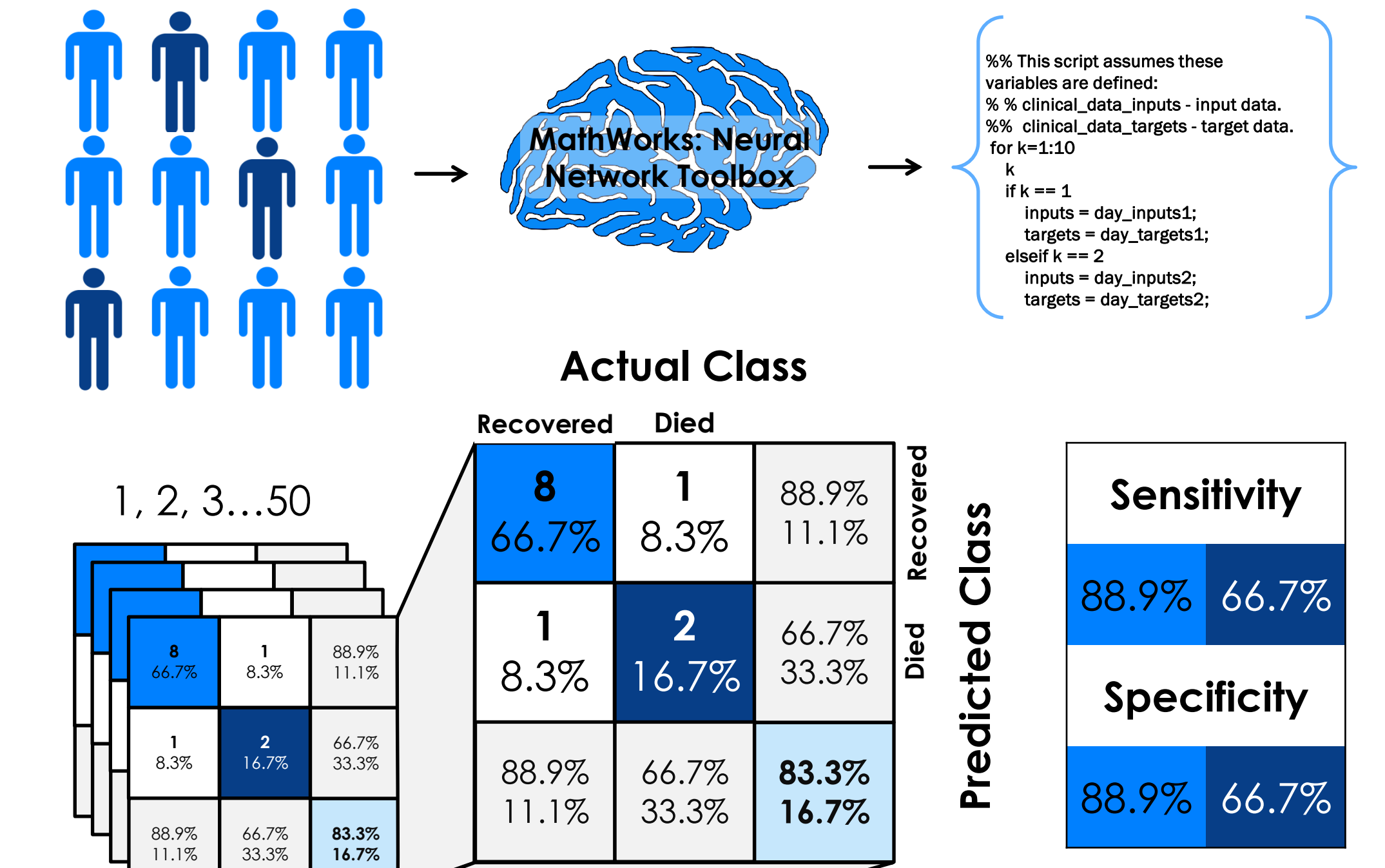


Figure 7. The accuracy of outcome based on daily vital signs recorded during the course of treatment. Points labeled *single* used data on that specific day of treatment whereas points labeled *recurring* incorporate the patients' vital signs on previous days of treatment.

Training Data Set



Conclusions

- SCNS, ENC, ARF, and HYPO are the four discrete variables with highest correlation with patient outcome.
- These four variables have high predictive powers with a sensitivity of 98.5% and 80.6% and a specificity of 95.7% and 92.6%, for recovered and deceased patients respectively.
- Older patients have higher case fatality rate. For instance, patients 21 and under have case fatality rate (CFR) of 8.3%, patients between the age of 45 and 54 have CFR of 26.9% and patients 65 and over have CFR of 36.4%
- Although it was thought to have a significant impact on CFR, there is no clear correlation between the days of fever prior to hospital presentation and a patient's outcome.
- A physician can more precisely determine outcome among patients who share similar clinical diagnosis using creatinine and lactate dehydrogenase lab results as shown in Figure 5.
- The difference in predicting outcome between single and recurring treatment data of patients vital signs is overall not significant, however there is a distinct peak on Day 9 for recurring data. In both cases there is a definite trend for increasing accuracy as treatment progresses.
- Zye is a powerful and precise tool for visually exploring and assessing multivariable data sets for correlations. For this Lassa fever case-study, Zye allows us to elegantly and intelligently generate hypotheses based on epidemiological data of Lassa fever patients.

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Andres Colubri. Without his guidance and support this project would not have been possible. I am grateful to Dr. Pardis Sabeti, all members of the Sabeti lab, and everyone at Fathom for this opportunity and valuable experience. This work was generously supported by the FAS Center for Systems Biology and HHMI Interdisciplinary Undergraduate Fellows Program.

References

- Asogun DA, Adomeh DI, Ehimuan J, Odiya I, Hass M, et al. (2012) Molecular Diagnostics for Lassa Fever at Irrua Specialist Teaching Hospital, Nigeria: Lessons Learnt from Two Years of Laboratory Operation. *PLoS Negl Trop Dis* 6(9): e1830.
- Center for Disease Control. 2013. Lassa Fever. [Online]. Available: <http://www.cdc.gov/ncidod/dvrd/spb/mnpages/dispages/lassa.htm> Accessed August 8, 2013.
- Khan SH, Goba A, Chu M, Roth C, et al. (2008) New opportunities for field research on the pathogenesis and treatment of Lassa fever. *Antiviral Res* 78: 103-115.
- McCormick JB, King IJ, Webb PA, Johnson KM, et al. (1987) A case-control study of the clinical diagnosis and course of Lassa fever. *J Infect Dis* 155: 445-455.
- McCormick JB, Webb PA, Krebs JW, Johnson KM, Smith ES (1987) A prospective study of the epidemiology and ecology of Lassa fever. *J Infect Dis* 155: 437-444.