

UTC Project Information	
Project Title	Big Data Methodologies for Simplifying Traffic Safety Analyses
University	TTI, VTTI, SDSU
Principal Investigator	Dominique Lord
PI Contact Information	d-lord@tamu.edu
Funding Source(s) and Amounts Provided (by each agency or organization)	Safe-D (Federal): \$115,637 Match (Non-Federal): \$178,253
Total Project Cost	\$293,890
Agency ID or Contract Number	Grant No: 69A3551747115 Project: 01-001
Start and End Dates	May 2017 - August 2018
Brief Description of Research Project	Data used for safety analyses have unique characteristics that are not found in other areas of research. For example, one important characteristic that has been recently documented in the literature is related to datasets that contained a large amount of zero responses. For such datasets, the number of sites where no crash is observed is so large that traditional distributions and regression models cannot be used efficiently. In another example, naturalistic data that are now available for transportation and safety analyses usually include terabytes of information, in which traditional statistical tools cannot be used efficiently for establishing relationships. So far, very little research has been devoted about how to handle such datasets and when specific analysis tools or statistical models could be used based on the characteristics of the data. The primary objectives of this research are to evaluate statistical and other related methods that could simplify the analysis of the unique attributes related to safety and transportation-related big data. The secondary objective is to present guidelines that can be used by researchers and practitioners for simplifying data analyses. The project can have a significant impact in safety and research related to big data.
Describe Implementation of Research Outcomes (or why not implemented) Place Any Photos Here	The anticipated results or outcomes are as follows: a review and discussion related to the limitations of data collection procedures and existing statistical methods; guidelines for determining when datasets could be changed for reducing the excess zero responses; guidelines for determining when multi-parameter models should be used over traditional models; paper(s) to be presented at a future Transportation Research Board Annual Meeting; paper(s) to be

	<p>submitted to peer-reviewed journals, such as Accident Analysis & Prevention and Analytic Methods in Accident Research. The results will be immediately implementable. The end users will be researchers and practitioners who handle different types of crash datasets.</p>
<p>Impacts/Benefits of Implementation (actual, not anticipated)</p>	<ul style="list-style-type: none"> • New tools or statistical methods for handling datasets with excess zero responses. • New tools or statistical methods for handling big data. • Better estimate of risk factors influencing crashes. • Advancements in the highway safety sciences.
<p>Web Links</p> <ul style="list-style-type: none"> • Reports • Project website 	<p>http://www.vtti.vt.edu/utc/safe-d/index.php/projects/big-data-methods-for-simplifying-traffic-safety-analyses/</p>