

# Big Data and Transportation Safety: Connecting the Dots

**Subasish Das, Ph.D.**, Associate Transportation Researcher, Texas A&M Transportation Institute (s-das@tti.tamu.edu)

**Greg P. Griffin, Ph.D.**, Assistant Professor, Dept. of URP, The University of Texas at San Antonio (gregpgriffin@utexas.edu)

Research conducted by



## Abstract

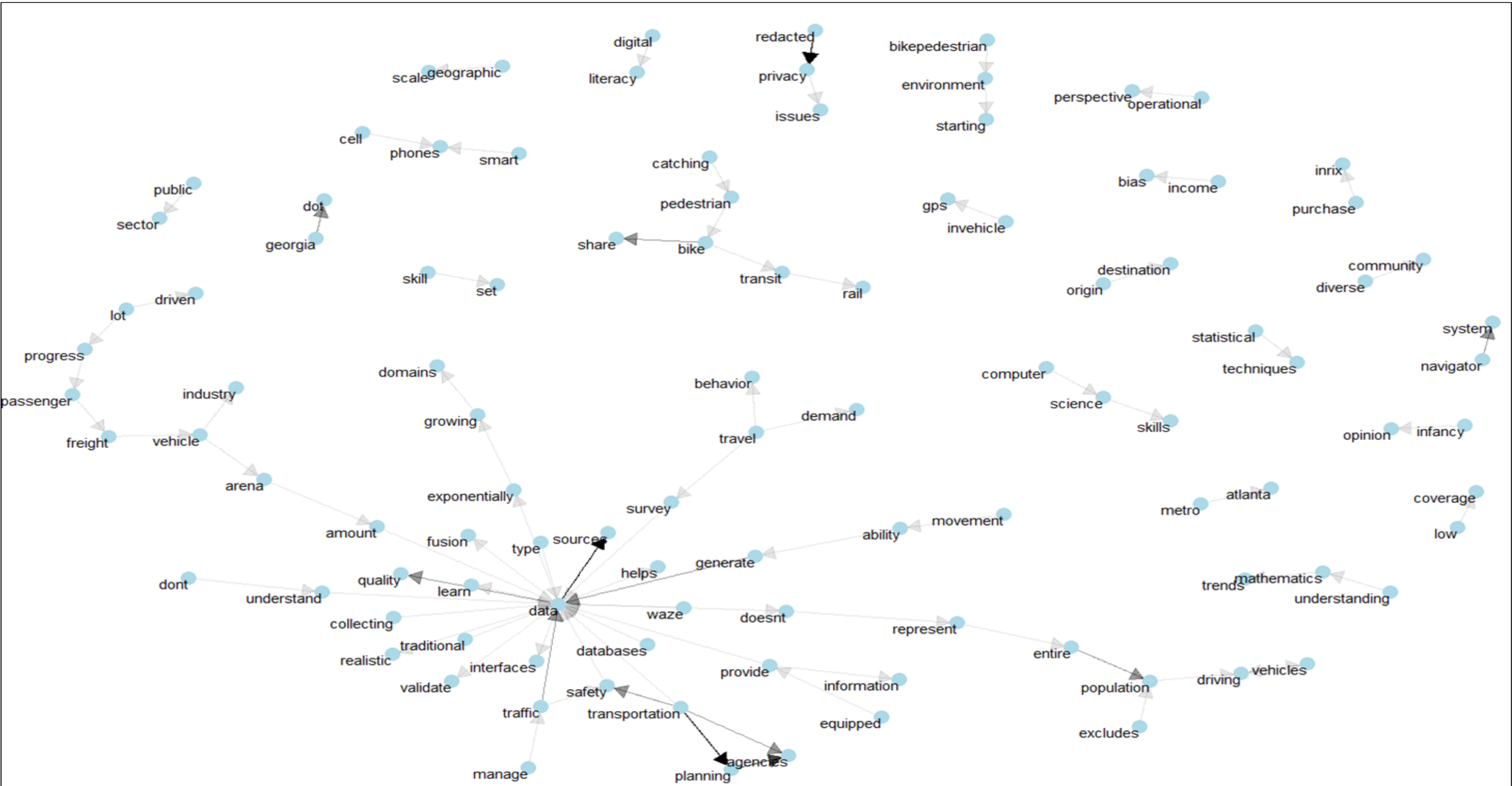
- Emerging big data resources and practices provide opportunities to improve transportation safety planning and outcomes.
- This study systematically reviewed both the sources of bias and approaches to mitigate them through a review of published studies and interviews with experts.
- The study includes a quantified analysis of topic frequency and an evaluation of the reliability of concepts through two independent trained coders.
- Applied natural language processing (NLP) tools to identify the trends in the unstructured textual contents.

## Survey Questions

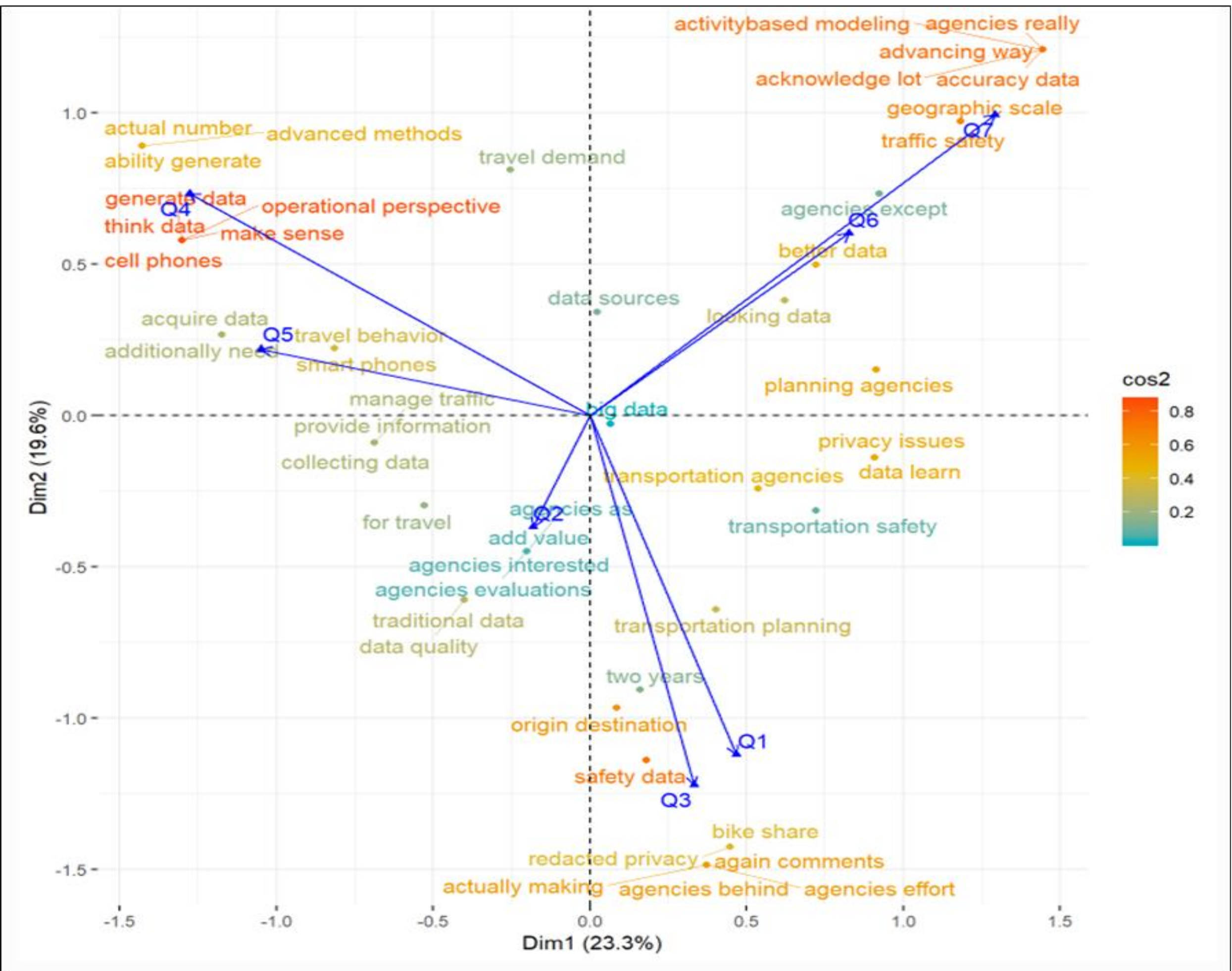
No.	Question
1	When did you start working with big data in transportation?
2	Why did your organization decide to use new sources of big data?
3	Has using big data helped improve transportation planning?
4	Are there ways that the data does not represent the entire population of interest in transportation planning?
5	How do you mitigate the impact of big data not representing the population?
6	Overall, has using big data has improved planning for transportation safety in your applications?
7	Is there anything else you would like to add?

MCA Parameters for Quadrant 1 (partial display of the complete table)

Bigrams	Coordinate		cos2		Contribution	
	Dim 1	Dim 2	Dim 1	Dim 2	Dim 1	Dim 2
data sources	0.023	0.343	0.001	0.125	0.001	0.342
looking data	0.622	0.381	0.229	0.086	0.957	0.426
better data	0.722	0.498	0.331	0.157	1.278	0.721
planning agencies	0.913	0.152	0.532	0.015	2.045	0.067
agencies except	0.923	0.732	0.074	0.047	2.110	1.576
geographic scale	1.184	0.971	0.398	0.268	3.472	2.768
traffic safety	1.184	0.971	0.398	0.268	3.472	2.768
accuracy data	1.445	1.209	0.451	0.316	5.173	4.295
acknowledge lot	1.445	1.209	0.451	0.316	5.173	4.295
activity-based modeling	1.445	1.209	0.451	0.316	5.173	4.295
advancing way	1.445	1.209	0.451	0.316	5.173	4.295
agencies really	1.445	1.209	0.451	0.316	5.173	4.295



Network Plot



MCA Plot

## Conclusions

- Four critical issues were identified about biases in big data, and how to mitigate:
- Keep transportation experts and public central in determining the right goals and metrics to evaluate transportation safety.
  - Develop new methods to relate big data to the total population needed for transportation safety.
  - Leverage big data to answer intractable questions.
  - Work ahead to transfer emerging knowledge to future problems.