

# Data Quality Control Example

## Comparison of Crash Dataset 2013 with 2012

David Brown [brown@cs.ua.edu](mailto:brown@cs.ua.edu)

August 22, 2014

The purpose of this document is to present an example of the use of IMPACT as a data quality control tool. The crash data will be used as an example, and a past year (2012) will be compared with the most current year (2013) in order to determine discrepancies between the 2013 and the 2012 datasets. Typically there are not dramatic changes in the values/distributions of these data elements from one year to the next. If major discrepancies are found, this will be useful for at least one of the following purposes:

1. To determine any major changes that are occurring in the nature of crashes that need to be addressed from a traffic safety point of view.
2. To determine if there are critical items that seem to be neglected by those doing the data entry in the most recent year;
3. To determine improvements that have been made in the data reporting.

These will be considered in separate sections below, each in order of the extent of the differences between the two years.

It is recommended that this analysis become a standard QC practice whenever there is a major addition of data to the system. This not only serves a quality control function, which is essential, but it also surfaces major changes that are critical to be detected at the highest level so that they can be subjected to further analyses.

### POTENTIAL TRAFFIC SAFETY EMERGING ISSUES

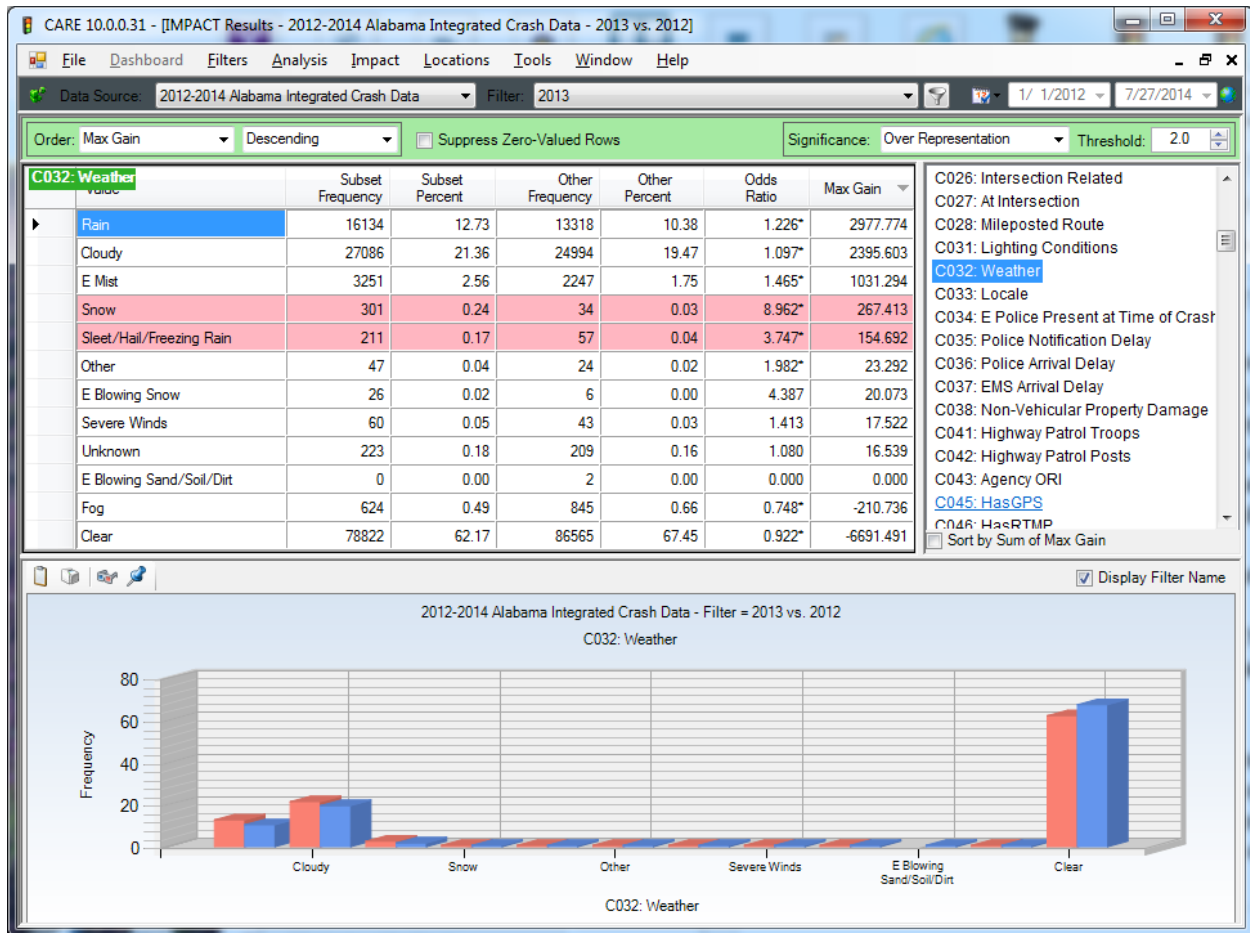
**C032 – Weather.** Clear weather was down from 67% to 62% of the reported crashes. All non-clear weather events were up sharply, especially Snow and Sleet/Hail/Freezing Rain. This was not due to a reporting, since this is a very reliable variable with less than 100 nulls in both years. This is no doubt an emerging safety issue and it should be given major consideration in formulating new traffic safety countermeasures. The citizens of Alabama are not generally experienced in handling these weather events effectively when on the road. See Display 1.

C403 – CU Roadway Condition – further supported the C032 results given above.

C225 – CU Citation Issued. The “E None Issued” category went from about 75% to 80%, a very significant increase that makes it clear that relatively fewer citations are being issued in connection with traffic collisions.

C501 – Vehicle 2 (V2) Type – generally this variable is not given too much consideration since V2 is considered to be the victim as opposed to the causal vehicle. However the number of “E Other Bus (Seats More than 15)” increased from 27 (0.02%) to 93 (0.07%) – and increase of about 3.5 times from what would be expected. This is a red flag that also needs looking into.

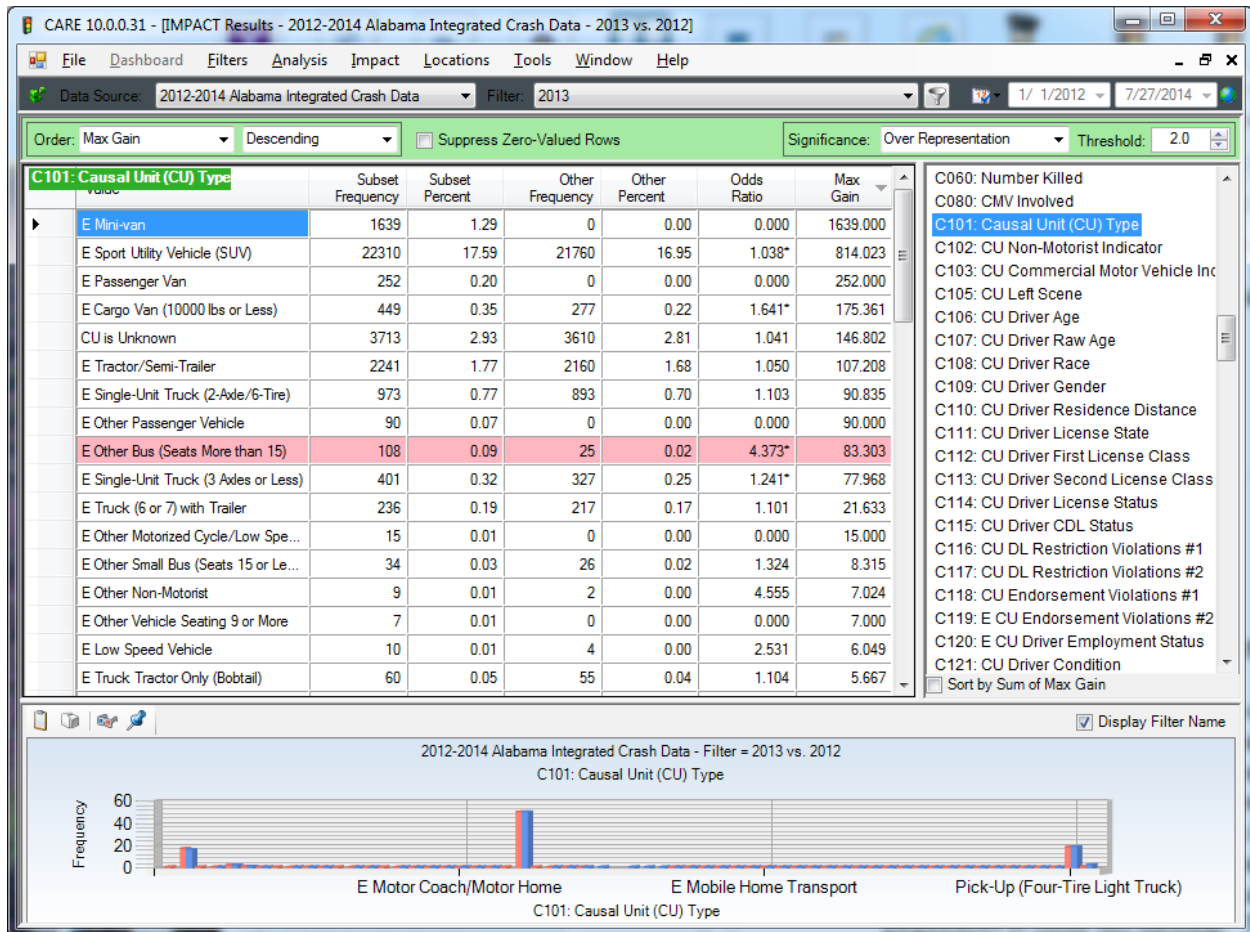
## Display 1. C032 Weather Comparison of 2013 (Red Bars) vs. 2012 (Blue Bars)



C101 – Causal Unit (CU) Type – Comparable to C501, there was a significant increase of the “E Other Bus (Seats More than 15)” from 25 in 2012 to 108 in 2013 (an increase of a factor of 4.4 times expected). This validated the fact that these types of vehicles need more intensive consideration. See Display 2.

C213 – CU Vehicle Usage – There was a 2.5 increase in the proportion of “E Vehicle Used as School Bus” from 42 (0.03%) to 103 (0.08%) a factor of 2.5 increase. Comparatively speaking these numbers are small, but, of course, special attention is required in any significant change involving school children. It is very difficult to not see that these last three variables must be related, and thus warrant considerable additional analysis.

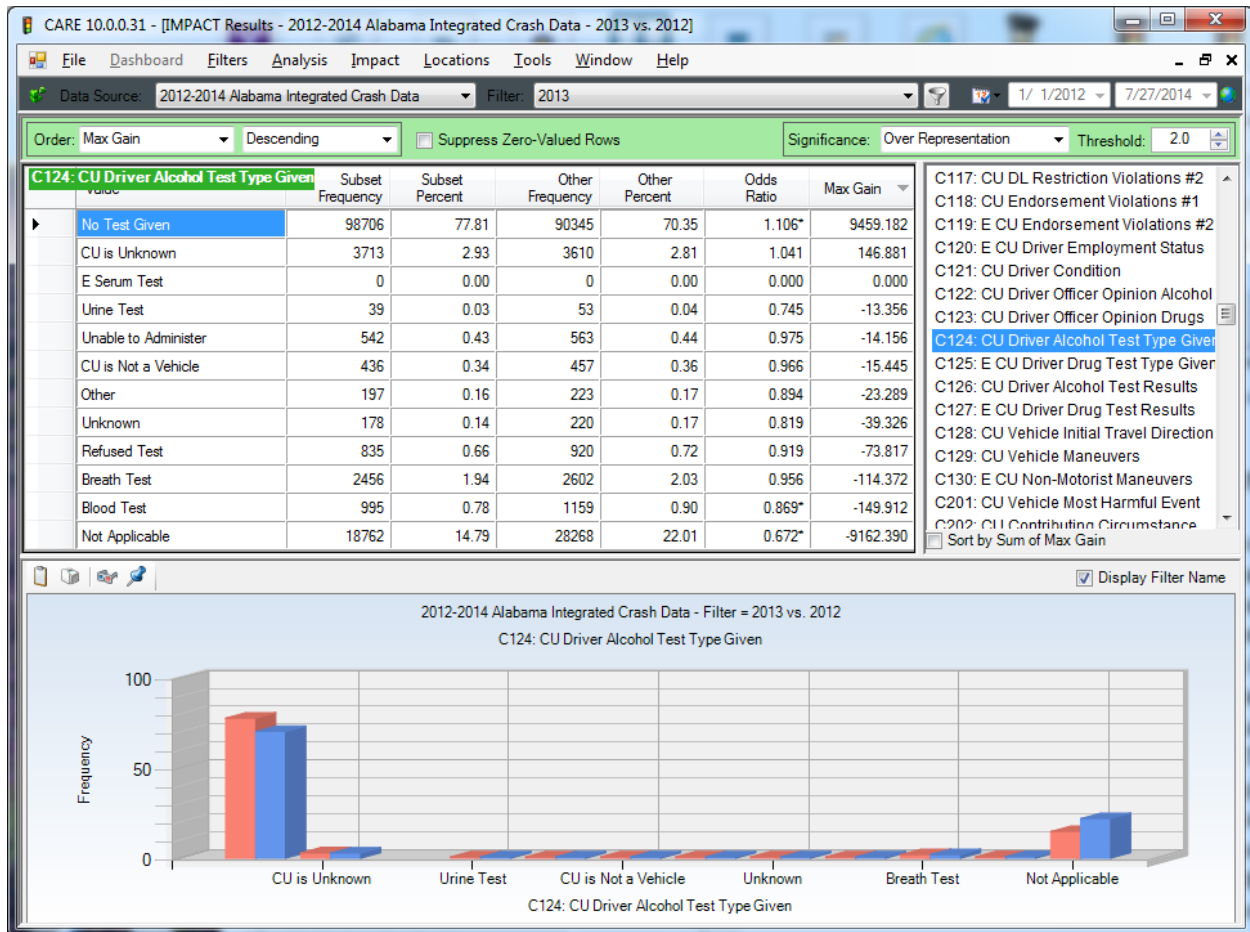
## Display 2. C101 Causal Unit Type Comparison of 2013 (Red Bars) vs. 2012 (Blue Bars)



## DETERIORATION IN REPORTING QUALITY

C124-C127 – there are four variables that provide an indication of the types of drug and alcohol tests given and their results. There has been an increase of at least 11% in the “No Test Given” category, and in the drug tests this was over 16% increase in the proportion. This is an alarming increase that indicates that officers are opting out of giving these tests, since there is no way that this could be any expected change in behavior from the driving public. One influencing factor here could be the mandatory breath test upon conviction of a DUI. Just how this would work is complex – it could be that some officers are reluctant to test because they do not want the perpetrators to suffer this hardship. This would not be the first time that this happened in Alabama – it was experienced during the early stages of the MADD insistence on jail time for all DUIs. On the other hand, some officers might be expecting that this law will go into effect and dramatically reduce the DUI problem without their intervention. It is recommended that this be further analyzed by jurisdiction and that survey contact be made in those agencies that have the most dramatic drops to determine the reason.

### Display 3. C124 Alcohol Test Type Comparison of 2013 (Red Bars) vs. 2012 (Blue Bars)



C216 – E CU Placard Status. The number of nulls in this variable increase proportionately by a factor of approximately 2.4. The explanation for this should be sought with the DPS CMV Unit.

C111 – CU Driver License State. The number of nulls in this variable increased over four-fold, from 1.69% in 2012 to 6.85% in 2012. This was accompanied by a reduction of “license state” results of Alabama from 79.5% to 74.0%. The conclusion must be that the problem involves all states licenses, including those within Alabama.

C511 – V2 Driver License State. The increase in nulls was even worse in the second vehicle drives – up nearly by a fact of six. This shows a consistent issue.

C112 (and C113) – CU Driver First (and Second) License Class. There was a proportionate increase in both of these for “Not an Alabama License” of about 50%. Both First and Second License Class increased from about 10% to 15%.

## IMPROVEMENT IN REPORTING QUALITY

C020 – E Distracted Driving – this is a new variable that had essentially no entries in 2012, so it is a major improvement in the completeness of getting information on our distracted driving issues.

C043 – Agency ORI – this needs to be given much more intense evaluation to track who is an who is not reporting. The largest increase, from 305 to 1492 (close to a factor of five proportionately) was seen by Jefferson County Sheriff’s Office, and they are to be commended. See Display 3.

C002 – City. This validated the discussion of C043 above, and this variable should be used in any further analyses. Recommended are cross-tabulations between these two variables and also by C001, County.

**Display 3. C043 Reporting Agency Comparison of 2013 (Red Bars) vs. 2012 (Blue Bars)**

