Part IV: REPORTS

Quarterly Progress Report Narrative for the First quarter.

Project Title: Crash Records Data Improvement Plan

Project Number: K9-13-18-05

Implementing Agency: DHSMV

Project Director: Maureen Johnson

Describe the subgrant activities that took place during the quarter. Attach newspaper clippings, press releases, photos and other items that document activities. Use additional sheets if necessary. The *Quarterly Progress Report of Performance Indicators* should be sent to the DOT Safety Office along with the narrative within 30 days of the end of each quarter.

An Analyst was hired Oct 15, who consulted with Federal Crash Data Improvement Project Technical Team regarding suggested cross-edits and best practices from other states. Though there were some universal cross-edits, it was highly recommended that Florida stakeholders establish areas of greatest concern that meet their business objects. The overall Florida census focused on meaningful feedback to law enforcement agencies regarding incomplete and inaccurate reports. Areas of interest were the accuracy of location data, person information and the quality of reporting commercial motor vehicle data.

One of the first goals was to establish an accuracy baseline. This was not completed during the first quarter. Accuracy could be measured at two levels: Level 1) crash reports in the production database that have logical inconsistences; Level 2) crash reports that didn't meet the minimum requirement of accuracy to be imported into the production database.

Currently the XML schema definitions, those standards provided vendors and law enforcement agencies, along with crash manual, show no cross-field edits. It was therefore assumed that level 2 accuracy would be permissive as it would allow logical inconsistences. It was determined that both level 1 and level 2 accuracy needed be established to understand the impact of adding cross-fields edits. It was assumed that initially any decrease in level 1 would result in an increase in level 2; and an understanding of both were necessary before determining an accuracy baseline.

Level 2 accuracy was measured by the total number of 2011 long form crash reports returned to law enforcement agencies for fatal errors (fatal error defined as a report rejected by the state). Seven percent (7.13%) were returned to agencies with this type of error, either by email notice (e-crash) or by mail (paper reports submitted for keypunch). There were 95 documented error types. Eighteen percent (18.29%) of fatal errors could not be documented as an error type (aborts).

A review of the 95 documented error types, showed several that could only be possible if there were cross-edits in effect. There are two types of cross-field edits: 1) if A then B is limited to a subset of attributes; 2) if A then B is required. Thirty-one percent of fatal errors were attributed to these two types of cross-field edits. These included time fields -- arrived time, cleared time, report date time – and others like non-motorist action prior to crash, non-motorist location code, date of birth, owners last name, etc. These cross-edits, or any of the load program code edits, had not been documented so impossible to compare against the current XML schema definitions. The analyst used the quarter to document the current load program edits to determine what cross-fields were already in place. This also provided an

opportunity to glean some insight in how effective cross-field edits would be.

The project goal is not just improved accuracy but completeness as well.

All fields on the crash report, where "Other" or "Unknown" was a possible attribute, were looked at to determine the frequency of use. The range here, when a null value was allowed – field not required -- was from approximately .6% (multiple fields) to 21% (Placarded).

Since there were already some cross-edits in place, providing an opportunity to judge effectiveness, another analysis was done on cross-edits of the type "if A then B is required." If a non-motorist was involved in the crash these fields were required: "non-motorist location at time of crash" and "non-motorist action prior to crash". In the case of "non-motorist location at time of crash" 31.25% entered "Other" or "Unknown"; in the case of "non-motorist action prior to crash" 27.97% entered "Other" or "Unknown."

It was concluded that the cross-edits of the type "if A then B" would be less effective if there wasn't an established mechanism for providing data quality feedback to the individual agency -- over relying upon "other " or "unknown" neither increased the accuracy or completeness of the crash report. Although no completeness baseline had yet been established it "non-motorist location at time of crash" and "non-motorist action prior to crash" would in all probability exceed the standard.

The analyst during the quarter had several meetings with the University of Florida Programmers who he would be working on to revise XML schema definitions. These meetings concentrated on the standardizing of Law Enforcement Agency names as they appear on a crash report. (They are currently free text – allowing any spelling.) Data for 2011 showed 6310 possible spellings for less than 400 law enforcement agencies. (It should be noted in 2012, with the growth of e-crash, this was down to 998 spellings). Once non- e-crash data was in the database it became, without a consistent or standardized Department name, difficult to correspond the crash data with the appropriate law enforcement agency. The importance of Standardized agency names was also discussed with Florida DOT for their efforts in locating crash locations.

During the quarter the analyst put forth two possible ways to create Department name standardization. The first was through the creation of an alias list. The second was to use the HSMV crash report numbers assigned to the agency by DHSMV to identify the law enforcement agency. These are currently under review.

QUARTERLY PROGRESS REPORT OF PERFORMANCE INDICATORS

for the First quarter.

Project Title: Crash Records Data Improvement Plan

Project Number: K9-13-18-05

Implementing Agency: DHSMV

Project Director: Maureen Johnson

Performance Indicators	Milestones Accomplished					
	Quarter Ending December 31	Quarter Ending March 31	Six-Month Totals	Quarter Ending June 30	Quarter Ending September 30	Project Totals
			0			0
			0			0
			0			0
			0			0
			0			0
			0			0
			0			0
			0			0
			0			0

Part IV: REPORTS

Quarterly Progress Report Narrative for the Second quarter.

Project Title: Crash Records Data Improvement Plan

Project Number: K9-13-18-05

Implementing Agency: Florida Department of Highway Safety and Motor Vehicles

Project Director: Maureen Johnson

Describe the subgrant activities that took place during the quarter. Attach newspaper clippings, press releases, photos and other items that document activities. Use additional sheets if necessary. The *Quarterly Progress Report of Performance Indicators* should be sent to the DOT Safety Office along with the narrative within 30 days of the end of each quarter.

The Analyst identified an additional 22 cross-field edits during the quarter. Seven single-field patterns or maximum length restrictions require updating in the XML schema, scheduled for the last quarter of the grant. Three rule types were established. Analyst conducted bi-weekly meetings with University of Florida Programmers. An alias list was created for standardizing the law enforcement agency names stored in the HSMV production database.

After looking at the complexity of a single accuracy or completeness baseline for crash reports it was determined a better approach would be baselines for each part of the crash report. Crash reports were separated into Event, Person and Vehicle, allowing flexibility that would encourage a more comprehensive strategy for accurate and complete crash data. Rather than a complicated formula for establishing a graded weight to each field, across multiple sections of the report, each major section would be evaluated separately.
This also allows the analysts some flexibility when establishing baselines, or reestablishing baselines, as the process is iterated.

For example, fields initially considered as a benchmark were "vehicle type" cross-referenced against "required endorsements". An analysis showed that for 40.27% of motorcyclists the endorsement field was null. A cross-field edit whereby if vehicle type = motorcycle then the required endorsement field will be required. This would effectively prevent null fields for the required endorsement and would have returned approximately 3600 crash reports back to law enforcement agencies.

Further investigation by the Analyst determined that this incompleteness of required endorsements was isolated to specific group of law enforcement agencies that shared a common vendor. A review of the vendor's data collection client uncovered a disassociation of motorcyclist with the required endorsement field. The analyst worked directly with the vendor to correct this problem, which was resolved mid-February.

A baseline for accuracy and completeness, using this benchmark as one consideration, could be shown to meet the overall performance goal of a 5% increase. To continue the iterative process a new baseline will be established with a different set of cross-field benchmarks used as performance measures.

The analyst worked on 1510 problems (aborts) related to e-crash submissions, as well as with one vendor for e-crash certification in addition to working with the vendor regarding the motorcycle endorsement errors.

The accuracy baseline1 error rate for "Event" was established at 29.62%. This was based on county name matching county code and Lighting conditions matching time of day. Of 296,890 reports submitted in 2011,

87,925 had errors in one or both of these fields.

A baseline1 for vehicle hasn't yet been established. A baseline2 for person hasn't yet been established.

QUARTERLY PROGRESS REPORT OF PERFORMANCE INDICATORS

for the Second quarter.

Project Title: Crash Records Data Improvement Plan

Project Number: K9-13-18-05

Implementing Agency: Florida Department of Highway Safety and Motor Vehicles

Project Director: Maureen Johnson

Performance Indicators	Milestones Accomplished					
	Quarter Ending December 31	Quarter Ending March 31	Six-Month Totals	Quarter Ending June 30	Quarter Ending September 30	Project Totals
Number of cross-edits identified	7	22	**29**			**7**
Number of single field pattern strengthen improvements	0	7	**7**			**0**
Number of updates to Crash Report Manual	0		**0**			**0**
			0			**0**
			0			**0**
			0			**0**
			0			**0**
			0			**0**
			0			**0**

Part IV: REPORTS

Quarterly Progress Report Narrative for the Third quarter.

Project Title: Crash Records Data Improvement Plan

Project Number: K9-13-18-05

Implementing Agency: Florida Department of Highway Safety and Motor Vehicles

Project Director: Maureen Johnson

Describe the subgrant activities that took place during the quarter. Attach newspaper clippings, press releases, photos and other items that document activities. Use additional sheets if necessary. The *Quarterly Progress Report of Performance Indicators* should be sent to the DOT Safety Office along with the narrative within 30 days of the end of each quarter.

The analyst reviewed the potential advantage and possible problems with implementing various cross-field edits identified in previous quarters with stakeholders. Consideration was given to increasing the accuracy and completeness of crash reports while not returning an excessive number to agencies. For instance, in 2011, if a cross-field edit such as "injury severity" matching the appropriate "form type" was imposed more than 18,000 additional crash reports would have been returned to law enforcement agencies.

There were concerns that if an excessive amount of crash reports were returned to LEAs the reports would not be not be corrected and returned to HSMV. It was determined that a system should be created that would track all returned reports and monthly notify each agency as to what reports were still outstanding (the error not fixed and the report resubmitted to HSMV) before imposing excessive edits.

Code changes were made to the load/error handling process that will greatly reduce or eliminate future aborts (crash reports that failed to be uploaded to HSMV and did not return an error message back to the LEA). While continuing to manually correct the 1510 aborted reports from the previous quarter and the 598 occurring during this quarter, the analyst identified and documented the types of problems causing aborts and code changes were made to the load/error handling process, so that aborts would be returned to LEAs as errors in the already established automated load/error reports flowing back to the agencies.

An emphasis remained on three initial focus areas for stakeholders: event (time and location), inconsistent reporting for alcohol involvement among drivers and non-motorists, and a more complete picture of commercial motor vehicles.

The accuracy baseline error rate for "Event" (time and location) was established at 29.62% during the second quarter. It was also decided during the second quarter that there would be separate baseline measurements for "Person" (inconsistent reporting for alcohol involvement among drivers and non-motorist) and "Vehicle" (a more complete picture of commercial motor vehicles). These were established during the third quarter. The cross-fields edits were proposed and will be implemented at the beginning of the fourth quarter.

Person – baseline 9.10%

In 2011, 7748 completed alcohol tests were reported (alcohol_test_results = '2')
for drivers and non-motorists.

A. 81.47% (or 6312) of them resulting in non-zero blood alcohol results

B. 18.06% (or 1399) of them reporting zero blood alcohol results
C. 00.48% (or 37) of them without missing(pending) blood alcohol results

Additionally, there were 1215 pending alcohol tests reported (alcohol_test_results = '1') for drivers and non-motorists.

D. .99% (or 12) of them resulting in non-zero blood alcohol results
E. 63.13% (or 767) of them resulting in zero blood alcohol results
F. 35.88% (or 436) of them with missing(pending) blood alcohol results

C , D and E are inconsistent results between the two fields.
Total inconsistent = 9.10% (37+12+767)/(7748+1215)

The following cross-field edits were suggested to reduce the number of inconsistent results:

1) Do not allow 'Pending' in alcohol_tested_results and values in blood_alcohol_content
2) Do not allow 'Completed' in alcohol_tested_results and no value in blood_alcohol_content

This will increase the number of pending alcohol results in the database, but will more accurately reflect the driving and non-motorist population.

Pending report s with zero (".00") blood alcohol results are not in line with the 18% completed reports with zero (".00") BAC, pointing to a problem with certain clients possibly pre-filling this data with zeros.

Vehicle – baseline 7.27%

In 2011, 613,636 vehicles that were marked as non-commercial vehicles, 15,293 vehicles were marked commercial and on 46,437 reports the flag was neither true nor false.

Of the 613,364 vehicles with the commercial vehicle flag set to False:
A. 99.96% (or 613,364) did not have a commercial motor vehicle configuration indicated
B. 00.04% (or 272) did have commercial motor vehicle configuration indicated

Of the vehicles with the commercial vehicle flag set to True:
C. 03.61% (or 552) did not have the commercial motor vehicle configuration indicated
D. 96.39% (or 14,741) did have the commercial motor vehicle configuration value indicated

Of the vehicles with the commercial vehicle flag unset :
E. 99.29% (or 46,106) did not have a commercial motor vehicle configuration indicated
F. 00.71% (or 331) did have a commercial motor vehicle configuration indicated

B, C and F are inconsistent results. Although the total percentage of inconsistent results across all vehicles is low (0.17% or (272+552+331)/(613636+15293+46437)), this is a large percentage (7.27% or (272+552+331)/(15293+272+331)) over just the commercial vehicles realm. Assuming crashes with

commercial motor vehicle configuration set represent commercial motor vehicles, the total commercial motor vehicles set to true should be 15292+603. If the configuration is set correctly, and these should be commercial motor vehicles, it will add 603 more vehicles into the commercial motor vehicle = 'T' category, for a percentage increase of 3.79% (603/(15292+603)).

The following cross-field edits were suggested to prevent inconsistencies:

1) If commercial vehicle flag is marked True, require commercial motor vehicle configuration be set
2) If commercial motor vehicle configuration is set, require commercial vehicle flag is marked True
3) Require commercial vehicle flag be either true or false

July 1, 2012 legislation changed the reporting requirements for long forms, adding commercial motor vehicles. This resulted in greater reporting of commercial motor vehicles and by doing so showing other potential need for cross-field edits that would not have been apparent from studying 2011 data. Hazardous Materials Released field being True ('2") and Hazardous Material Placard field being null was most notable.

Year	Placarded is null	Spilled='2'	PCT err
2011	6	574	1.05%
2012	19	2915	0.65%
2013	64	255	25.10%

+

Proposed edit:

1) If Hazardous Materials Released = True, then value for Hazardous Material Placard required.

As in the second quarter with motorcycle endorsement, if the analyst is able to track this sudden high rate of error back to a group of agencies utilizing the same data collection software, he can work directly with that party to improve their data collection. The analyst does not include the field in the calculation for improved accuracy since baseline numbers are based on 2011 data.

QUARTERLY PROGRESS REPORT OF PERFORMANCE INDICATORS

for the Third quarter.

Project Title: Crash Records Data Improvement Plan

Project Number: K9-13-18-05

Implementing Agency: Florida Department of Highway Safety and Motor Vehicles

Project Director: Maureen Johnson

Performance Indicators	Milestones Accomplished					
	Quarter Ending December 31	Quarter Ending March 31	Six-Month Totals	Quarter Ending June 30	Quarter Ending September 30	Project Totals
Number of cross-edits identified	7	22	**29**	0		**29**
Number of single field pattern strengthen improvements	0	7	**7**	0		**7**
Number of updates to Crash Report Manual	0	0	**0**	0		**0**
			0			**0**
			0			**0**
			0			**0**
			0			**0**
			0			**0**
			0			**0**

Part IV: REPORTS

Quarterly Progress Report Narrative for the Fourth quarter.

Project Title: Crash Records Data Improvement Plan

Project Number: K9-13-18-05

Implementing Agency: Florida Department of Highway Safety and Motor Vehicles

Project Director: Maureen Johnson

Describe the subgrant activities that took place during the quarter. Attach newspaper clippings, press releases, photos and other items that document activities. Use additional sheets if necessary. The *Quarterly Progress Report of Performance Indicators* should be sent to the DOT Safety Office along with the narrative within 30 days of the end of each quarter.

The proposed cross-field edits from the previous quarter were implemented in HSMV load program as warnings. Previous to implementing these edits the analyst had been tracking the impact for the separate sections of the crash report – Event, Person, Vehicle – where accuracy baselines had been established. This allowed the analyst to work with agencies avoiding any serious disruption with their submitting process. By working with the agency directly the agencies were able to adjust to significantly improve accuracy across the three sections of the crash report. This was done without increasing the number of reports returned to the agencies in the automated load/error reports received by agencies after each submission. As accuracy continues to improve, and fall within a reasonable threshold for returned reports, the warnings will become hard edits during the second year of the grant, not allowing reports to load into the HSMV database until they comply .

The crash report was separated into three sections as noted – Event, Person, Vehicle -- allowing each individual section to have its own accuracy error rate, which can also be expressed as a logical inconsistency. During the grant period:

- The accuracy error rate for EVENT decreased from the 2011 established baseline by 58.94%, moving from an error rate of 29.62% down to 12.16% (see I below)
- The logical inconsistencies for PERSON decreased from the 2011 established baseline by 27.58% , moving from an error rate of 9.10% down to 6.50% (see II below)
- The logical inconsistencies for VEHICLE decreased from the 2011 established baseline by 37.61%, moving from an error rate of 6.97%* down to 4.39% (see III below).

**(This baseline was adjusted from the quarter before where inadvertently FHP short form crash reports were included in the calculation. This moved the VEHICLE baseline from 7.27% to 6.97%)*

A second goal, excessive use of "Unknown" or "Other", to bypass providing tangible values for required fields, was also reviewed. This was secondary to accuracy error rate, and directly related to logical inconsistencies. For instance, with the PERSON section, an Alcohol Test Result field cannot be "Unknown" if there is a value in the Blood Alcohol Content field. Likewise, for the Vehicle section as the percent of logical inconsistencies becomes lower the use of "Unknown" and "Other" should also decrease.

PERSON

In 2011, 27,054 crash reports, with a reported value for Blood Alcohol Content, had a value of "Unknown" for

Alcohol Test Results. This was out of 35,535 crash reports where Alcohol Test Results were populated. "Unknown" percent was 76.56%. ("Other" was not applicable to these fields.)

In 2013 YTD (Oct 10), 15,787 crash reports, with a reported value for Blood Alcohol Content, had a value of "Unknown" for Alcohol Test Results. This was out of 21,097 crash reports where Alcohol Test Results were populated. "Unknown" percent was 74.83%.

This demonstrated a reduction rate of 2.25% for the use of "Unknown" in the Alcohol Test Results where there was a value for Blood Alcohol Content. The analyst has identified agencies where crash reports are being defaulted to some variation of "0", rather than leaving the field blank for Blood Alcohol Content when the Alcohol Test Results is "Unknown." (Having identified these agencies, it is expected, in the second year of the grant, to see a significant reduction in the percent of reports where these fields are logically inconsistent.)

VEHICLE
In 2011, it was determined that 14,852 vehicles should have been reported as a commercial motor vehicle (see III below). Of those 1,740 had a commercial motor vehicle configuration as "Unknown" and another 672 had a commercial motor vehicle configuration as "Other." This showed a combined use rate of 16.24% for these terms for commercial motor vehicles.

In 2013, it was determined (YTD) 13,486 vehicles should have been reported as a commercial motor vehicle. Of those 810 had a commercial motor vehicle configuration as "Unknown" and another 528 had a commercial motor vehicle configuration as "Other." This showed a combined use rate of 9.92%.

This demonstrated a reduction rate of 38.91% for the use of "Unknown" and "Other" in the commercial motor vehicle configuration field.

I.
EVENT BASELINE – 29.62%
Event baseline was based on county code matching county literal, and time of day having a logical relationship to lighting conditions.

In 2011 there were 296,890 reports submitted, 87,925 had errors in one or both of these fields. Event accuracy error rate was established at 29.62%.

EVENT 2013
In 2013 YTD (Oct 10) there were 290,168 long forms submitted, 35,296 had errors in one or both of these fields. Event accuracy error rate was 12.16%.

EVENT RESULTS (end of year one grant period)
The accuracy error rate for EVENT decreased from 2011 established baseline by 58.94% during the grant period.

II.
PERSON BASELINE – 9.10%

Person baseline was based on blood alcohol content have a logically consistent relationship with the status of the alcohol test result field.

In 2011 there were 7,748 alcohol test results marked as "completed" for drivers and non-motorists.

A. 81.47% (or 6312) of them resulting in non-zero blood alcohol results
B. 18.06% (or 1399) of them reporting zero blood alcohol results
C. 00.48% (or 37) of them without missing(pending) blood alcohol results

Additionally, there were 1,215 "pending" alcohol tests reported for drivers and non-motorists.

D. .99% (or 12) of them resulting in non-zero blood alcohol results
E. 63.13% (or 767) of them resulting in zero blood alcohol results
F. 35.88% (or 436) of them with missing(pending) blood alcohol results

C , D and E are inconsistent results between the two fields. Logical inconsistent baseline established for Person was 9.10% (or (37+12+767)/(7748+1215)).

PERSON 2013
In 2013 YTD (Oct 10) there were 5,011 alcohol test results marked as "completed" for drivers and non-motorists.

A. 81.97 % (or 4,108) of them resulting in non-zero blood alcohol results
B. 16.18 % (or 811) of them reporting zero blood alcohol results
C. 1.83 % (or 92) of them without missing(pending) blood alcohol results

Additionally, there were 1,571 "pending" alcohol tests reported for drivers and non-motorists.

D. 2.10 % (or 33) of them resulting in non-zero blood alcohol results
E. 19.66 % (or 309) of them resulting in zero blood alcohol results
F. 78.23% (or 1,229) of them with missing(pending) blood alcohol results

C , D and E are inconsistent results between the two fields. Logical inconsistencies for Person was 6.59%. (92+33+309)/(5011+1571)

PERSON RESULTS (end of year one grant period)
The logical inconsistencies for PERSON decreased by 27.58% (or 9.10 – 6.59/9.10) during the grant period.

III.
VEHICLE BASELINE – 6.97%
Vehicle baseline was based on commercial motor vehicle flag being consistent with the commercial motor vehicle configuration field.

In 2011, 518,122 vehicles were marked as non-commercial vehicles, 14,288 vehicles were marked commercial and on 33,140 reports the flag was neither true nor false.

Of the 518,122 vehicles with the commercial vehicle flag set to False:
A. 99.94% (or 517,852) did not have a commercial motor vehicle configuration indicated
B. 00.05% (or 270) did have commercial motor vehicle configuration indicated

Of the 14,288 vehicles with the commercial vehicle flag set to True:
C. 03.30% (or 472) did not have the commercial motor vehicle configuration indicated
D. 97.01% (or 13,861) did have the commercial motor vehicle configuration value indicated

Of the 33,140 vehicles with the commercial vehicle flag unset :
E. 99.11% (or 32,846) did not have a commercial motor vehicle configuration indicated
F. 00.88% (or 294) did have a commercial motor vehicle configuration indicated

B, C and F are inconsistent results. The total percentage of inconsistent results across all vehicles was low 0.18% (or (270+472+294)/(518122+14288+33140)), this is a large percentage (6.97% or (270+472+294)/(14288+270+294)) over just the commercial vehicles realm. Assuming crashes with commercial motor vehicle configuration set represent commercial motor vehicles, the total commercial motor vehicles set to true should have been 14288+564. If the configuration was set correctly, and these should have been commercial motor vehicles, it would have added 564 more vehicles into the commercial motor vehicle True category, for a percentage increase of 3.79% (564/(14288+564)).

VEHICLE 2013

In 2013 YTD (Oct 10) 476,452 vehicles were marked as non-commercial vehicles (False), 13,270 vehicles were marked commercial (True) and on 62,134 reports the flag was neither true nor false.

Of the 476,452 vehicles with the commercial vehicle flag set to False:
A. 99.95% (or 476,236) did not have a commercial motor vehicle configuration indicated
B. 00.045% (or 216) did have commercial motor vehicle configuration indicated

Of the 13,270 vehicles with the commercial vehicle flag set to True:
C. 02.84% (or 377) did not have the commercial motor vehicle configuration indicated
D. 97.15% (or 12,893) did have the commercial motor vehicle configuration value indicated

Of the 62,134 vehicles with the commercial vehicle flag unset :
E. 100% (or 62,134) did not have a commercial motor vehicle configuration indicated
F. 0% (or 0) did have a commercial motor vehicle configuration indicated

B, C and F are inconsistent results. The total percentage of inconsistent results across all vehicles was 0.10% (or (216+377+0)/(476452+12270+62134)). The percentage over just the commercial vehicles was 4.39% or (216+377)/(13270+216+0)).

Assuming crashes with commercial motor vehicle configuration set represent commercial motor vehicles, the total commercial motor vehicles set to true should be 13270+216. If the configuration was set correctly, and these should be commercial motor vehicles, it would add 216 more vehicles into the commercial motor vehicle "True" category, for a percentage increase of 1.60% (216/(13270+216)).

VEHICLE RESULTS (end of year one grant period)
The logical inconsistencies for VEHICLE decreased by 37.61% (or 6.97 - 4.39/6.97) during the grant period.

QUARTERLY PROGRESS REPORT OF PERFORMANCE INDICATORS

for the Fourth quarter.

Project Title: Crash Records Data Improvement Plan

Project Number: K9-13-18-05

Implementing Agency: Florida Department of Highway Safety and Motor Vehicles

Project Director: Maureen Johnson

Performance Indicators	Milestones Accomplished					
	Quarter Ending December 31	Quarter Ending March 31	Six-Month Totals	Quarter Ending June 30	Quarter Ending September 30	Project Totals
Number of cross-edits identified	7	22	**29**	0	4	**33**
Number of single field pattern strengthen improvements	0	7	**7**	0	0	**7**
Number of updates to Crash Report Manual	0	0	**0**	0	20	**20**
			0			**0**
			0			**0**
			0			**0**
			0			**0**
			0			**0**
			0			**0**

Final Narrative Report

Project Title: Crash Records Data Improvement Plan

Project Number: K9-13-18-05

Implementing Agency: Department of Highway Safety and Motor Vehicles

Project Director: Maureen Johnson

The following is a chronological narrative history of the above listed project in accordance with *Part V: Acceptance and Agreement, Conditions of Agreement, 1. Reports.* This report is an accurate accounting of the project performance and accomplishments. Attach additional sheets as needed.

During the first quarter, Tom Austin was hired to serve as the analyst on this project and he consulted with State and Federal stakeholders regarding suggested cross-field edits and best practices from other states. He also reviewed and analyzed 95 documented error types and identifed 7 cross-field edits.

During the second quarter an additional 22 cross-field edits were identified, and 7 single field patterns were identified. Bi-weekly meetings with University of Florida Programmers were held. An alias list for law enforcement agency names was created, and the accuracy baseline 1 error rate for "Event" was established.

During the third quarter, discussions with stakeholders regarding the potential advantages and possible problems with implementing the various cross-field edits previously identified took place.Code changes were made to the load/error handling process. Baseline measurements for "Person" and "Vehicle" were established.

During the fourth quarter,cross-field edits were implemented as warnings. Four additional cross-filed edits were identified, and 20 updates were made to the Crash Report Manual.

Throughout the project, the analyst worked closely with law enforcement agencies and software vendors to resolve errors.

See attached report for in-depth evaluation of objectives and the rationale and methodologies used.

CRASH RECORDS DATA IMPROVEMENT PLAN K9-13-18-05, AQS22
EVALUATION

Objectives for the crash data quality improvement project were to increase accuracy and completeness of long form crash reports and develop productive communications with law enforcement agencies to improve the quality of submissions. During the grant period long form crash report accuracy increased by 20.53%; and completeness increased by 27.57%.

Accuracy and completeness was to be based on selected fields where cross-field edits could be put in place to tighten submission requirements. Both federal and state stakeholders were consulted to determine what first round of edits would best serve their needs as users of the data, with a long term goal that this would be an iterative process.

Though there were some universal cross-field edits identified, it was highly recommended that Florida stakeholders' business needs take precedent. Areas of interest were the accuracy of location data, being about to report the day and the time of day crashes occurred, reporting and tracking the blood alcohol content for drivers and non-motorist where there were alcohol test results indicated, and a more complete reporting of commercial motor vehicle data.

OBJECTIVE ONE

- Establish a baseline for accuracy. To increase the accuracy rate of crash data submitted to the Department by 5%.

Based on 2011 long form crash report data a baseline accuracy rate was established at 73.42%. This was determined from potential cross-field edits that would immediately address stakeholder's interests; and after a complete review and documenting of the already existing cross-field edits in place during the first quarter of the grant.

The crash report was separated into three sections: Event, Person and Vehicle.

Event:

- County code matches county literal
- Time of day had a logical relationship to lighting conditions

Person:

- Blood alcohol content has a logically consistent relationship with the status of the alcohol test result field

Vehicle:

- Commercial motor vehicle flag was logically consistent with the commercial motor vehicle configuration field

For the accuracy baseline each section was weighted based on its potential population of crash reports. As such EVENT was 85.49%, PERSON was 10.23% and VEHICLE was 4.27% and multiplied against their accuracy grade: 70.38%, 90.90% and 93.03%, respectively. The weighted scores were used to determine the accuracy baseline percentage.

OBJECTIVE TWO

- Establish a baseline for completeness . Increase the completness rate of crash data submitted to the Department by 5%.

Based on 2011 long form crash report data a completeness rate was established at 44.32%. This was to address use of "Unknown" and "Other" as values in fields where a more tangible value should have been known or use of the terms appeared excessive; and files left blank that should have been required given other information on the crash report.

The crash report was not separated into the same three sections as for accuracy. Even though EVENT lighting condition did have the potential use of "Unknown" or "Other" in 2011 these values were reported less than 1% of the time, demonstrating that "Unknown" or "Other" was not being overused. However, a Florida stakeholder was seeing a high amount of nulls for the required endorsement field when clearly the completion of this field was warranted.

The three areas used for completeness score: Person, Required Endorsements and Vehicle.

PERSON

- If blood alcohol content had a value then the status of the alcohol test result field could not be "Unknown" ("Other" was not applicable to these fields)

REQUIRED ENDORSEMENTS

- If the type of vehicle is a motorcycle then the required endorsement field cannot be null

VEHICLE

- When commercial motor vehicle flag was set to true how high was the percentage of reports where commercial motor vehicle configuration field was either "Unknown" or "Other"

For the completeness baseline each section was weighted based on its potential population of crash reports. As such PERSON was 59.55%, REQUIRED ENDORSEMENTS was 15.55% and VEHICLE was 24.89% and multiplied against their completeness grade: 23.44%, 61.33% and 83.76%, respectively. The weighted scores were used to determine the completeness baseline percentage.

BEYOND OBJECTIVES ONE AND TWO

Though fields were specified for performance measures that did not preclude the analyst from identify others in need of edits during the grant period.

For instance, on July 1, 2012 legislation changed the reporting requirements for long forms, adding commercial motor vehicles. This resulted in greater reporting of commercial motor vehicles and by doing so showing other potential need for cross-field edits that would not have been apparent from studying 2011 data. Hazardous Materials Released field being True ('2") and Hazardous Material Placard field being null was most notable.

Year	Placarded is null	Spilled='2'	PCT err
2011	6	574	1.05%
2012	19	2915	0.65%
2013	64	255	25.10%

Edit:

- If Hazardous Materials Released = True, then value for Hazardous Material Placard is required.

In all, 29 edits were identified and/or documented. These were either already in place but undocumented or proposed.

IMPACT OF PROPOSED EDITS

The analyst further assessed the impact of proposed edits. In 2011, Seven percent (7.13%) of long form crash reports were returned to law enforcement agencies for not meeting minimum edit criteria. Edits included time fields -- arrived time, cleared time, report date time – and others like if a non-motorist was involved in the crash then the fields "non-motorist action prior to crash" and "non-motorist location code" were required.

Accuracy error rates for the proposed edits were established, which showed the percentage of long form crash reports that would be returned to Law Enforcement Agencies if the proposed edits were implemented into the current load/error return program.

- The 2011 the accuracy error rate, which could also be expressed as a logical inconsistency, for EVENT was 29.62%.
- The 2011 logical inconsistency, or accuracy error rate, for PERSON was 9.10%
- The 2001 logical inconsistency, or accuracy error rate, for VEHICLE was 6.97%

This meant for EVENT alone another 29.62% (or 87,925 out of 296,890) crash reports submitted would have been returned to Law Enforcement Agencies to be corrected and resubmitted.

BALANCE

It was already suspected that 2011 long form crash reports* were underreported. In 2010 there were 314,417 submitted to the state; in 2011 there were 296,890; in 2012 there were 345,039.

**(These are raw Long form crash reports and should not to be confused with long form crash reports that meet the criteria for statistical reporting as found in publications such as Crash Facts.)*

The state did not want to return tens of thousands of crash reports and then have to chase after them to be corrected and resubmitted. In fact, the state wanted to bring the volume of long form crash reports back up to expected numbers. So, a balance needed to be reached, that of having to increase accuracy and completeness without returning excessive amount of reports or discouraging law enforcement agencies from submitting all long form crash reports.

Instead of increasing the load/error return rate to agencies, the analyst worked directly with those agencies and vendors that had high percentages of inaccurate reports as defined by performance measures or incomplete reports also defined by the performance measures. (Even after edits were added to HSMV load program (fourth quarter), these were flagged as warnings rather than returns. This avoided any serious disruption with their submitting process.)

For example, after further investigation of REQUIRED ENDORSEMENTS (as a measurement for completeness) the analyst determined that this incompleteness was isolated to a specific group of law enforcement agencies that shared a common e-crash vendor. A review of the vendor's data collection client uncovered a disassociation of motorcyclist with the required endorsement field. The analyst worked directly with the vendor to have their software client corrected.

Another example was the accuracy of EVENT. A large portion of logical inconsistencies between time of day and lighting condition was also traced back to several agencies under contract with the same vendor (a different vendor than in the first example). For some agencies the time component in a crash_date_time field was not being exported to HSMV. Again this was resolved with the analyst working directly with the software vendor.

It should be noted that the volume for crash reports for 2013 (YTD – Oct.17) was 298,132. This showed that as accuracy and completes scores increased the volume of long form crash reports compared to 2011 had also significantly increased.

As accuracy continues to improve, and fall within a reasonable threshold for returned reports, the warnings will become hard edits during the second year of the grant, not allowing reports to load into the HSMV database until they comply.

YEAR ONE RESULTS

OBJECTIVE ONE

Increase the accuracy rate of crash data submitted to the Department by 5%.

Accuracy rate from 2011 baseline was increased by 20.53%

- 2011 baseline accuracy 73.42%. (See 2011 Accuracy Rate)
- 2013 accuracy rate 88.50%. (See 2013 Accuracy Rate)

2011 Accuracy Rate:
Accuracy grade 100% minus EVENT error rate; 100% minus PERSON error rate; 100% minus VEHICLE error rate (or 70.38%, 90.90%, 93.03%). Weighed based on population: 296,890, 35,535, 14,852, respectively. Event value was 85.49% applied to accuracy score, Person 10.23%, vehicle 4.27%. And these values multiplied against accuracy grade and then added together to determine accuracy baseline (or, 60.16 + 9.29 + 3.97 = 73.42).

2013 Accuracy Rate:
Accuracy grade 100% minus EVENT error rate; 100% minus PERSON error rate; 100% minus VEHICLE error rate (or 87.84%, 93.50%, 95.61%). Weighed based on population: 290,168, 21,097, 13,486, respectively. Event value was 89.35% applied to accuracy score, Person 6.49%, vehicle 4.15%. And these values multiplied against accuracy grade and then added together to determine accuracy rate (or, 78.84 + 6.06 + 3.96 = 88.50).

OBJECTIVE TWO

Increase the completeness rate of crash data submitted to the Department by 5%.

Completeness rate from 2011 baseline was increased by 27.57%

- 2011 baseline completeness 44.32%. (See 2011 Completeness Rate)
- 2013 accuracy rate 56.54%. (See 2013 Completeness Rate)

2011 Completeness Rate:
Completeness grade 100% minus PERSON error rate; 100% minus REQUIRED ENDORSEMENTS error rate; 100% minus VEHICLE error rate (or 23.44%, 61.33%, 83.76%). Weighed based on population:

35,535, 9,278, 14,852 respectively. Person value was 59.55% applied to accuracy score, required endorsements 15.55% , vehicle 24.89%. And these values multiplied against accuracy grade and then added together to determine accuracy baseline (or, 13.95 + 9.53 + 20.84 = 44.32).

2013 Completeness Rate:
Completeness grade 100% minus PERSON error rate; 100% minus REQUIRED ENDORSEMENTS error rate; 100% minus VEHICLE error rate (or 25.17%, 83.72%, 90.08%). Weighed based on population: 21,097, 7,788, 13,468 respectively. Person value was 49.81% applied to accuracy score, required endorsements 18.38%, vehicle 31.79%. And these values multiplied against accuracy grade and then added together to determine accuracy rate (or, 12.53 + 15.38 + 28.63 = 56.54).

ERROR RATES FOR ACCURACY

- The accuracy error rate for EVENT decreased from the 2011 established baseline by 58.94%, moving from an error rate of 29.62% down to 12.16% (see I below)
- The logical inconsistencies for PERSON decreased from the 2011 established baseline by 27.58%, moving from an error rate of 9.10% down to 6.50% (see II below)
- The logical inconsistencies for VEHICLE decreased from the 2011 established baseline by 37.61%, moving from an error rate of 6.97%* down to 4.39% (see III below).

ERROR RATES FOR COMPLETENESS

PERSON ERROR RATE

In 2011, 27,054 crash reports, with a reported value for Blood Alcohol Content, had a value of "Unknown" for Alcohol Test Results. This was out of 35,535 crash reports where Alcohol Test Results were populated. "Unknown" percent was 76.56% ("Other" was not applicable to these fields.)

In 2013 YTD (Oct 10), 15,787 crash reports, with a reported value for Blood Alcohol Content, had a value of "Unknown" for Alcohol Test Results. This was out of 21,097 crash reports where Alcohol Test Results were populated. "Unknown" percent was 74.83%

REQUIRED ENDORSEMENTS ERROR RATE

In 2011, 3,588 drivers of motorcycles showed no value for the required endorsement field on the crash report. This was out of 9,278 motorcycle drivers reported on long form crash reports. A null valve accounted for 38.67%

In 2013 YTD (Oct 10) , 1,268 drivers of motorcycles showed no value for the required endorsement field on the crash report. This was out of 7,788 motorcycle drivers reported on long form crash reports. A null valve accounted for 16.28%

VEHICLE ERROR RATE

In 2011, it was determined that 14,852 vehicles should have been reported as a commercial motor vehicle (see III below). Of those 1,740 had a commercial motor vehicle configuration as "Unknown" and

another 672 had a commercial motor vehicle configuration as "Other." This showed a combined use rate of 16.24% for these terms for commercial motor vehicles.

In 2013, it was determined (YTD) 13,486 vehicles should have been reported as a commercial motor vehicle. Of those 810 had a commercial motor vehicle configuration as "Unknown" and another 528 had a commercial motor vehicle configuration as "Other." This showed a combined use rate of 9.92%.

I.

EVENT ERROR RATE BASELINE – 29.62%

Event baseline was based on county code matching county literal, and time of day having a logical relationship to lighting conditions.

In 2011 there were 296,890 reports submitted, 87,925 had errors in one or both of these fields. Event accuracy error rate was established at 29.62%.

EVENT 2013

In 2013 YTD (Oct 10) there were 290,168 long forms submitted, 35,296 had errors in one or both of these fields. Event accuracy error rate was 12.16%.

EVENT RESULTS (end of year one grant period)

The accuracy error rate for EVENT decreased from 2011 established baseline by 58.94% during the grant period.

II.

PERSON ERROR RATE BASELINE – 9.10%

Person baseline was based on blood alcohol content have a logically consistent relationship with the status of the alcohol test result field.

In 2011 there were 7,748 alcohol test results marked as "completed" for drivers and non-motorists.

A. 81.47% (or 6312) of them resulting in non-zero blood alcohol results
B. 18.06% (or 1399) of them reporting zero blood alcohol results
C. 00.48% (or 37) of them without missing(pending) blood alcohol results

Additionally, there were 1,215 "pending" alcohol tests reported for drivers and non-motorists.

D. .99% (or 12) of them resulting in non-zero blood alcohol results
E. 63.13% (or 767) of them resulting in zero blood alcohol results
F. 35.88% (or 436) of them with missing(pending) blood alcohol results

C, D and E are inconsistent results between the two fields. Logical inconsistent baseline established for Person was 9.10% (or (37+12+767)/(7748+1215)).

PERSON 2013

In 2013 YTD (Oct 10) there were 5,011 alcohol test results marked as "completed" for drivers and non-motorists.

A. 81.97 % (or 4,108) of them resulting in non-zero blood alcohol results
B. 16.18 % (or 811) of them reporting zero blood alcohol results
C. 1.83 % (or 92) of them without missing(pending) blood alcohol results

Additionally, there were 1,571 "pending" alcohol tests reported for drivers and non-motorists.

D. 2.10 % (or 33) of them resulting in non-zero blood alcohol results
E. 19.66 % (or 309) of them resulting in zero blood alcohol results
F. 78.23% (or 1,229) of them with missing(pending) blood alcohol results

C, D and E are inconsistent results between the two fields. Logical inconsistencies for Person was 6.59%. (92+33+309)/(5011+1571)

PERSON RESULTS (end of year one grant period)

The logical inconsistencies for PERSON decreased by 27.58% (or 9.10 – 6.59/9.10) during the grant period.

III.

VEHICLE ERROR RATE BASELINE – 6.97%

Vehicle baseline was based on commercial motor vehicle flag being consistent with the commercial motor vehicle configuration field.

In 2011, 518,122 vehicles were marked as non-commercial vehicles, 14,288 vehicles were marked commercial and on 33,140 reports the flag was neither true nor false.

Of the 518,122 vehicles with the commercial vehicle flag set to False:

A. 99.94% (or 517,852) did not have a commercial motor vehicle configuration indicated

B. 00.05% (or 270) did have commercial motor vehicle configuration indicated

Of the 14,288 vehicles with the commercial vehicle flag set to True:

C. 03.30% (or 472) did not have the commercial motor vehicle configuration indicated
D. 97.01% (or 13,861) did have the commercial motor vehicle configuration value indicated

Of the 33,140 vehicles with the commercial vehicle flag unset :

E. 99.11% (or 32,846) did not have a commercial motor vehicle configuration indicated
F. 00.88% (or 294) did have a commercial motor vehicle configuration indicated

B, C and F are inconsistent results. The total percentage of inconsistent results across all vehicles was low 0.18% (or (270+472+294)/(518122+14288+33140)); however, this is a large percentage (6.97% or (270+472+294)/(14288+270+294)) over just the commercial vehicles realm. Assuming crashes with commercial motor vehicle configuration set represent commercial motor vehicles, the total commercial motor vehicles set to true should have been 14288+564. If the configuration was set correctly, and these should have been commercial motor vehicles, it would have added 564 more vehicles into the commercial motor vehicle True category, for a percentage increase of 3.79% (564/(14288+564)).

VEHICLE 2013

In 2013 YTD (Oct 10) 476,452 vehicles were marked as non-commercial vehicles (False), 13,270 vehicles were marked commercial (True) and on 62,134 reports the flag was neither true nor false.

Of the 476,452 vehicles with the commercial vehicle flag set to False:

A. 99.95% (or 476,236) did not have a commercial motor vehicle configuration indicated
B. 00.045% (or 216) did have commercial motor vehicle configuration indicated

Of the 13,270 vehicles with the commercial vehicle flag set to True:

C. 02.84% (or 377) did not have the commercial motor vehicle configuration indicated
D. 97.15% (or 12,893) did have the commercial motor vehicle configuration value indicated

Of the 62,134 vehicles with the commercial vehicle flag unset :

E. 100% (or 62,134) did not have a commercial motor vehicle configuration indicated
F. 0% (or 0) did have a commercial motor vehicle configuration indicated

B, C and F are inconsistent results. The total percentage of inconsistent results across all vehicles was 0.10% (or (216+377+0)/(476452+12270+62134)). The percentage over just the commercial vehicles was 4.39% or (216+377)/(13270+216+0)).

Assuming crashes with commercial motor vehicle configuration set represent commercial motor vehicles, the total commercial motor vehicles set to true should be 13270+216. If the configuration was set correctly, and these should be commercial motor vehicles, it would add 216 more vehicles into the commercial motor vehicle "True" category, for a percentage increase of 1.60% (216/(13270+216)).

VEHICLE RESULTS (end of year one grant period)

The logical inconsistencies for VEHICLE decreased by 37.61% (or 6.97 - 4.39/6.97) during the grant period.