

1. Will this presentation be posted publicly on the NHTSA website?

A: Yes.

2. Is this information in a report?

A: NHTSA is documenting all the details of the sample design underlying CISS and CRSS and will publish these in the form of a report.

3. Need a reminder of what SSU, PJ, MOS, TSU abbreviations mean - thanks!

A: Please see attached Glossary of terms towards the end of this document.

4. Will you be disseminating the new methodology changes on the NHTSA website?

A: All the details of the new design will be published in the form of a report. In addition, guidance will be issued for proper analysis of the data. These publications will be available on NHTSA's website.

5. What will happen to all of the CDS and GES data already obtained? Will it be incorporated in CISS and CRSS somehow?

A: NHTSA will continue to use the CDS and GES data already obtained. They will not be 'incorporated' into CISS or CRSS. However, we will issue guidance on how one might combine data from the two systems appropriately (e.g., a researcher may want to combine CDS and CISS data to get a reasonable sample size for analysis).

6. For the MY sampling, how do you choose when there are multiple vehicles involved in a collision?

A: The latest (newest) model year vehicle that was towed is used for listing and sampling purposes.

7. By selecting the CISS PSUs based on having 5+ fatalities/year, aren't we designing a system that is skewed towards the more severe crashes that occur and not necessarily capturing an accurate picture of what is happening? How did you arrive at the 5 fatal crash count criteria?

A: This 5+ fatality/year condition was just one factor used to form the PSUs, not select them. In other words, when we formed the PSUs, we tried to ensure the PSUs are large enough with a high probability (90% chance) that there are at least 5 fatalities per year in each PSU. When we selected the PSU sample however, all PSUs, large or small, have a

chance to be selected. We are not selecting the CISS PSUs based on having 5+ fatalities per year.

The 5 fatalities were used as one factor in PSU formation based on the assumption that each CISS data collection site will have at least one technician. From our past experience, we know one technician can finish about 100 crash investigations per year. Our goal of having fatal crashes comprising 5% of our sample translates into 5 fatal crashes per year for one technician. Therefore, in order to support at least one technician's workload, the 'at least 5 fatal crash per year' criteria was used in the PSU formation.

8. As crash avoidance technology experiences higher market penetration and theoretically reduces the number of crashes and the severity of crashes, are we going to have trouble maintaining PSUs with 5+ fatal crashes/year? And as this happens, won't CISS be farther and farther from an accurate picture of what is actually happening on roads? I think it's great to know what the more severe crashes are, but I think it is also important to have an accurate assessment of how many crashes happen each year and how severe they are. I am concerned that this system may not capture that.

A: The 5+ fatal crashes/year condition used in PSU formation is based on current PSU fatal crash counts. For many reasons, the PSU fatal crash count distribution will likely change in the future.

The impact of those changes is that the estimates will be less efficient (larger variance) in the future but still unbiased. NHTSA is considering weighting adjustment using other external information to keep future estimates efficient, relevant and current.

9. Is Fan saying that 5% of the sampled CISS crashes will be fatalities? If 5000 cases are sampled, is he saying that only 250 fatal crashes will be investigated? If not, about how many fatal crashes will be investigated in CISS in a given year.

A: Yes, if 5,000 crashes are sampled in a year, the target number of fatal crashes to be sampled will be 250 (based on the 5% target sample distribution).

10. When will the current NASS CDS and GES Sample Collection end and the CRSS and CISS begin?

A: CDS ends with the 2015 data collection year. For CISS, NHTSA plans to field a 5-site Pilot in 2015 and have 24 sites up and running for Phase 1 by end of 2016. When CISS has 24 sites fully operating, we expect to collect 4,000 cases per year.

GES ends with the 2015 data collection year. For CRSS, NHTSA plans full implementation in 2016 in 60 sites generating about 50,000 cases annually.

11. Will NHTSA provide new software tools for the end-user to query these relational databases other than the XML case viewer tools they currently offer which has limited and restricted search capability?

A: Output of the public files will include continuance of the SAS data file to maintain consistency, but will also include other output formats of data. Data users will have the ability to view common reports and data requests in the future in a set of web-based filterable views, which will also allow for extracts to be produced in formats that can be used in such tools as Excel by allowing for common delimited formats to be generated. Users will also have the ability to download collections of formatted data on a pre-defined commonly used subject in a comma delimited format to allow a user that does not have SAS the ability to still analyze the subset of the data.

NCSA is also working on presenting data such as fatality datasets in a graphical format using GIS presentation and business intelligence toolsets, which will allow the public user a more flexible and user friendly interface when reviewing views of interest.

12. What improvements in the estimates of the national numbers of fatalities will the revisions in the CRSS accomplish compared with the problems of GES in terms of estimates of the total number of fatalities (i.e., the discrepancy in the fatality estimates compared with FARS)? Same question for the CISS?

A: A major area of focus is on weight calibration. Studies are underway to perform a series of adjustments using auxiliary information that will be collected by NHTSA. For example, estimates of fatal crashes from CRSS will be consistent with those reported from FARS (a census of all fatal crashes).

13. Will the agency implement a process for downloading the images associated with the data in a similar manner to which the data itself can be downloaded? The images are often critical to understanding cases of interest when the data is incomplete. The current process of using the online case query tool is too cumbersome for efficient analysis.

A: Due to the volume and complexity of object storage and its relationship to case management, the images have to be closely tied to case viewers. NHTSA is reviewing options for meeting a variety of business uses. At this time, we are looking at redesigning the case viewers to be more user friendly, which appears to meet the most common business requirements around image viewing. However, over time NHTSA will continue to evaluate options that are most usable yet affordable.

14. When will the new data and associated weights be available?

A: Under our current plans, 2016 CRSS data will be released in late 2017 with the corresponding weights and analytic guidance. CISS is more complicated because it is being phased in over several years. NHTSA will make a determination on the appropriate usage of the data collected during the pilot phase after case data are available to evaluate. For CISS data collected in 2016 during the phased implementation of the PSUs, NHTSA plans to release it in late 2017 with weights and the appropriate analytic guidance.

15. People sometimes use NASS/CDS and get wildly unreliable estimates. Will NHTSA be providing guidance on the minimum size of particular categories needed to provide reliable estimates? For example, should people use CISS to estimate fatal crashes in certain categories or are those too rare in in CISS?

A: NHTSA will provide guidance on variance estimation using CISS or CRSS data so that data users can estimate the standard errors of their domain estimates. It is up to the data user to decide whether the associated precision is acceptable or not.

16. When will the CDS/GES stop being produced and replaced by the complete new databases?

A: Please see answer to question 10.

17. Is there any consideration to improving the Police reported severity scale?

A: In the 4th Edition of MMUCC, guidelines have been provided for adoption of a better set of definitions for the Injury Status KABCO attributes which NHTSA encourages states to adopt.

18. How much effort would it take to increase the focus on fatal accidents and decrease the sampling of no injury and injured cases in the CISS. why is there not more emphasis in the sampling scheme to focus more on fatal crashes

A: Please see answer to question 25.

19. For the CISS crashes involving heavy trucks, will the heavy truck information be collected?

A: CISS will collect general information on the heavy trucks, but not the detailed vehicle assessments done on passenger vehicles. In NHTSA's report to Congress, it is

recommended that special studies be used for more in-depth data collection on special populations, such as heavy trucks.

20. Can you please provide more information on the crash avoidance systems/elements that will be included in CISS? Thank you.

A: CISS will be collecting 8 pre-crash elements that describe what the vehicle was doing just prior to the crash. These elements such as Distraction, critical event, crash type, and pre first harmful event maneuver sequence are especially useful in determining scenarios where crash avoidance technologies could have been beneficial. Sources are scene inspection, vehicle inspection, and interview. In addition, CISS will be collecting information on the presence and activation of 5 crash avoidance technologies based on vehicle inspection and interviews.

FCW without auto braking

FCW with auto braking

LDW without lane keeping

LDW with lane keeping

Blind spot detection

21. Will CISS and CRSS be linked to FARS? This would be quite helpful.

A: This is an interesting suggestion and NHTSA will explore options when CISS and CRSS data are available.

22. Regarding precision, will there be any results published comparing census counts (FARS) and CISS estimates for fatalities. Also, Injury rates.

A: NHTSA is studying the effects and methods of calibration. NHTSA expects to calibrate CISS weights using external information so that estimates for a population of interest (e.g., fatal crashes in which at least one passenger vehicle was towed) will be consistent across multiple data systems.

23. You mentioned that during NHTSA's assessment of data needs the community said they wanted CDS's scope expanded to include more motorcycle, pedestrian, and large truck crashes, but all you did to the CDS scope was to remove the "towed due to damage" criterion. How do you plan to address this need?

A: Motorcycle and pedestrian are relatively rare populations. CISS is not designed to capture rare populations, so it would be inefficient to use CISS to collect in-depth motorcycle and pedestrian data at all CISS sites. However, NHTSA recommends designing and conducting special studies when funding becomes available to do special studies for pedestrians, large trucks, motorcycles and other populations of interest.. These designs will be tailored and optimized to get these crashes of.

24. I noticed that there is great interest in fatal crashes and concern that not enough will be sampled. My suggestion is to increase the SCI budget so that more fatal crash investigations can be conducted. This appears to be a more cost-effective measure than trying to increase the number of PSUs in CISS.

A: Please see answer to question 25.

25. What is the agency philosophy that suggests that 30% of CISS resources should be allocated to understanding no injury crashes while only 5% of CISS resources are being allocated to fatal crashes?

A: Although continued improvements in vehicle crashworthiness will still help reduce fatalities and injuries, NHTSA believes that the greatest gains in highway safety in the coming years will result from broad-scale application of crash avoidance technologies. Crash avoidance technologies not only would reduce fatalities and injuries, these technologies also would mitigate roadway congestion by preventing crashes. In order to evaluate crash avoidance technologies and crashes in general, the agency needs in-depth data for a crash database that is representative of the real-world crash population. Based on 2013 GES, minor injury and property-damage-only crashes comprised more than 85 percent of the police-reported crashes. Including these crashes in the CISS would allow the agency to establish a complete delta-v and injury profile, understand vehicle interactions and driver reactions in these crashes and examine the variability of crash avoidance performance with crash severity. Consequently, this will enable the agency to improve the benefit estimates for crash avoidance technologies and promulgate data-driven rulemakings on crash avoidance countermeasures. Furthermore, understanding these minor crashes also is critical to the research of autonomous vehicle technologies.

26. What can you say about the precision of the estimates from the new system as compared to those from the old system?

A: As we have seen from the optimum sample allocation results, the number of PSUs is the main driver of the precision. Under the current budget, the new surveys have the similar PSU sample sizes as the existing surveys. However, the new samples are selected with the updated frame information. In addition, the CISS PAR scalability will allow us to increase the number of useful cases dramatically. Use of auxiliary information at weighting stage also has the potential to reduce the variance dramatically. Currently NHTSA is studying the effects and methods of calibration. We expect calibration using external information will also further improve the precision.

Therefore we expect the new surveys to produce estimates with precision at least as good, if not better, than estimates from the existing surveys for the major domain estimates.

27. As a longtime NASS data user, I've noticed that some of the case weights are really high. Will this problem still be present in the new system? Why can't I just use the data without applying the weights, anyway?

A: In CDS, 'high' or large weights are a consequence of limited sample size. This is further exacerbated by the oversampling of severe crashes and therefore under-sampling of non-severe crashes. With the CDS sample size being lower than ever due to budget constraints, large weights have become inevitable. In the new system, we plan to take the following measures to control and reduce the large weights:

1. NHTSA will closely monitor the sampling process in CISS and CRSS. Once potential large weights are identified in the sample selection process, we'll adjust the sampling parameters to prevent the large weights.
2. NHTSA is considering smoothing the large weights through a calibration process to reduce the large weights.

Because of the oversampling of the severe crashes and the under-sampling of the non-severe crashes, the sample distribution is completely different from the population distribution. Therefore, un-weighted estimates using sample data can be severely biased for population parameters. However, under certain circumstances, such as some clinical studies, it is possible that the sample design become "non-informative" or "ignorable". In these situations, weighting may not be needed. Users should refer to related literature for further information.

Glossary

CDS: Crashworthiness Data System

CISS: Crash Investigation Sampling System (replacing CDS)

CRSS: Crash Report Sampling System (replacing GES)

GAO: Government Accountability Office

GES: General Estimate System

LMY: Late Model Year

MAP-21: Moving Ahead for Progress in the 21st Century Act

MMY: Medium Model Year

MOS: Measure of Size

NASS: National Automotive Sampling System

OMY: Older Model Year

PAR: Police Crash Report

PJ: Police Jurisdiction

PPS: Probability Proportional to Size

PSU: Primary Sampling Unit (1st Stage)

SPS: Sequential Poisson Sampling

SSU: Secondary Sampling Unit (2nd Stage)

TSU: Tertiary Sampling Unit (3rd Stage)