

Considerations for Analysis of the Common Core of Data (CCD)

Slide 1 of 26

Title Slide: Considerations for Analysis of the Common Core of Data (CCD)

Slide 2 of 26

This module describes the analytic considerations that should be kept in mind when using the Common Core of Data (or, CCD). It describes the importance of understanding the details when reviewing the data. It also describes universe data file inconsistency, and how to address data issues such as missing data, not applicable data, zero counts, misreported data, and misinterpreted data associated with the CCD.

Slide 3 of 26

When using CCD data, it is critical to pay attention to the various data details to ensure that data are being used properly. This includes reviewing the definitions, using the correct level of data, and noting the footnotes on the tables.

Slide 4 of 26

When using CCD data, be sure to understand what the data represent. It is easy to misinterpret output without a strong understanding of the data being used. The best way to understand the data is to read the documentation. For each variable, there is a detailed description. This is particularly important when reviewing the fiscal variables since the types of expenditures can vary.

Slide 5 of 26

As an example we'll look at the free lunch eligibility counts on the CCD data file. CCD reports the number of reported students that are eligible for free or reduced-price lunch.

Many people use this as a measure of poverty, since they assume that it shows the number of students who participate in the program.

The difference between "participating" and "eligible" could result in errors, however. In addition, the different methods used to determine eligibility could cause variation in the counts.

Slide 6 of 26

CCD data are typically reported at three levels – school, district, and state.

Make sure that all of the data come from the same level, if possible.

Slide 7 of 26

Here's an example, a researcher wants to create a table with the percentage of English-Language Learner (ELL) students by state.

Considerations for Analysis of the Common Core of Data (CCD)

To calculate the percentage, the user will need the total number of students. But which level? School? District/LEA? State?

The enrollment counts at the district level is the correct answer since the ELL counts are reported at the district (or Local Education Agency) level. The correct level can be determined by reviewing the documentation. It is important that the same level be used for both sets of numbers.

Slide 8 of 26

There are multiple tables available on the NCES website that include CCD data.

These include the tables from the CCD Reports, CCD table library, and the Digest of Education.

When comparing the tables to the CCD data files be sure to look at the footnotes and table title to make sure the differences between the two are taken into consideration.

Slide 9 of 26

Some of the differences between the data files and the tables may be due to data being from only one level, or the schools or districts are excluded for certain reasons (for example, only active schools or only districts with enrollment counts, or where U.S. totals are different than the universe, which includes outlying areas).

Slide 10 of 26

The CCD is used as a sampling frame for other data collections like the Civil Rights Data Collection (or, CRDC) and the Schools and Staffing Survey (or, SASS). However, the universes may not always match. For example, three schools on the CCD may be combined on SASS. In addition, the districts on the CCD may not be included on the CRDC collection, or vice-versa. It is important to keep these differences in mind as they may cause problems when comparing counts between the files.

Slide 11 of 26

The two universes between the CCD fiscal data files and nonfiscal data files may not always match. While the nonfiscal district file is used as the base for the fiscal (F-33) file, differences can still occur. For example, the nonfiscal collection may contain information about charter schools, but the fiscal collection may not be available. In addition, the enrollment counts for the two files may differ due to various adjustments being made to match the provided fiscal data on the file.

Slide 12 of 26

Within the data files, it's important to understand the differences between the terms missing data, not applicable data, zero counts, misreported data, and misinterpreted data. Specifically, missing data occur when the element exists but could not be reported. Not applicable occurs when the element does not apply to that case (for example, 12th graders in an elementary school). A count of zero indicates that the

Considerations for Analysis of the Common Core of Data (CCD)

element could exist but doesn't. Misreported data are those data where NCES questions the quality of the data and uses various methodologies to determine the extent to which they are useable. And, misinterpreted data is a problem if the data are not being used in the way they are intended. Each of these issues to consider will be discussed in the following slides.

Slide 13 of 26

Missing data occur when the element exists but could not be reported. This usually occurs if the reporting agency is not able to report the data, or if the data were not reported to NCES in time.

Slide 14 of 26

Here are a few examples of missing data. The number of teachers at a Kindergarten through 6th grade school could not be split into Kindergarten and Elementary teachers. The number of students who are eligible for free lunch could not be reported in time to be used on the data file. The state could not report which schools are magnet schools, but the state does allow for magnet schools.

Slide 15 of 26

In order to compare the free lunch eligibility of two districts, researchers will need to aggregate the school level enrollment and free lunch eligibility for each. Here we can see two districts, District A and District B, with their total free lunch eligibility, total enrollment, and percent eligible for free lunch.

When aggregating the free lunch and enrollment counts for District A, we can see that it has 120 students out of 300 that are eligible for free lunch. In other words, 40% of the total enrollment for District A is eligible for free lunch.

Similarly, District B has 60 students out of 300, or 20%, who are eligible for free lunch.

Slide 16 of 26

If we include missing data, however, a different finding may result.

In this example, District B had two schools out of the three where the free lunch data are missing. However, the enrollment count for all three schools was included in the aggregation. Without that information, it is unclear if the two districts should be compared. The total free lunch eligible count for District B could not be determined since a majority of the schools have missing data, so the percentage of students who are eligible for free lunch cannot (and should not) be calculated.

Slide 17 of 26

Researchers will also need to manage the missing data. On the CCD files, NCES indicates that the data element is missing by coding it either -1 or an M. For the data tools and reports, the aggregate totals at the state and district cannot be reported if 10%

Considerations for Analysis of the Common Core of Data (CCD)

of the data are missing (15% for U.S. totals). Missing data at the state level are imputed if the data are missing, but not at the lower levels.

Slide 18 of 26

On the data files, not applicable occurs when the element does not apply to that case. One example may be 12th graders in an elementary school.

A count of zero indicates that the element could exist but doesn't. For example, a small school with grades kindergarten through 12th grade that does not have any 8th graders.

Slide 19 of 26

Misreported data are those data where NCES questions the quality of the data and uses various methodologies to determine the extent to which they are useable. In order to verify the integrity of the data, three tests can be implemented:

- Comparing data from one year to the prior;
- Comparing data across data files, such as comparing the aggregate of school enrollment to the agency or state level counts;
- Data reported that are not consistent with the definitions, such as a school with students but no teachers.

Some issues can be explained and some data may need to be suppressed.

Slide 20 of 26

Some examples of situations with misreported data include the following:

The enrollment of a school changes from 50 to 300 but the teacher count stays the same. In this case, NCES may suppress the number of students. Or a Vocational school that offers grades K-12. In this case, NCES may change the grades to match the enrollment counts at the school.

Slide 21 of 26

In order to manage misreported data, NCES contacts the states to determine if the problem can be resolved (corrected or explained). Explanations are included as state notes within the documentation. If the data cannot be corrected or explained, NCES will indicate that the data element does not meet NCES standards by suppressing it and coding it to -9.

Slide 22 of 26

The CCD data can be misinterpreted if the data are not being used in the way they are intended. This can be an issue since the data could be used to indicate a situation that does not necessarily exist.

Some examples of misinterpreted data are presented on the next slide.

Considerations for Analysis of the Common Core of Data (CCD)

Slide 23 of 26

One type of misinterpretation is making incorrect assumptions. For example, it is incorrect to assume that eligibility for the free lunch program is the same thing as household poverty.

It is also incorrect to assume that the graduation rate is 100% of students minus the dropout rate. The Average Freshman Graduation Rate and the Adjusted Cohort Graduation Rate are better indicators of the graduation rate. In addition, it's critical to take into consideration the difference between event dropout (such as, students who dropout in a single year) and status dropout (such as, the percentage of high school dropouts among persons 16 through 24 years old) rates.

It is also incorrect to assume that "full time equivalency" are equal to head counts.

Another type of misinterpretation is an incorrect comparison. Sometimes researchers may make incorrect comparisons of data as well. Some examples of incorrect comparisons are noted here. Current expenditures are not the same as total expenditures. District and school enrollment are not always comparable. United States, reporting states, and totals from the data file do not indicate the same information. And, race counts between years where the number of categories change from five to seven.

Slide 24 of 26

Other issues to be aware of include not being careful when aggregating data (including missing data), schools with zero students may be shared-time schools, using the correct school year (noting school year vs. fiscal year), or improperly recoding (for example, agency type).

Slide 25 of 26

There are several ways to avoid misinterpreting data. First, make sure to read the documentation, especially the state notes, as important exceptions will be noted there. Second, researchers must have an accurate understanding of what the data do (and do not) represent. Third, review how the data were collected. Finally, verify that the coding was done correctly including handling missing, not applicable, suppressions, changes in definitions, and so on.

Slide 26 of 26

This module has described the analytic considerations that should be kept in mind when using the Common Core of Data (or, CCD) including attention to detail and universe data file inconsistency. It also described the analytic considerations associated with data issues such as missing data, not applicable data, zero counts, misreported data, and misinterpreted data associated with the CCD.

Considerations for Analysis of the Common Core of Data (CCD)

Important resources that have been provided throughout the module are summarized here for your reference.

You may now exit the module.