

ACACA data release and presentation guidelines

Introduction

ACACA agencies have rich and correct data sets about students' participation and achievement in senior secondary studies. These data sets can provide useful illumination of the educational challenges and successes in the senior secondary sector. Poorly used, however, the data sets can obscure and mislead, potentially harming individuals and organisations.

The data sets involved are distinct from those expected in well-designed research studies:

- the data are complete in the sense that results are held for all students involved in senior secondary studies
- the data are not complete in the sense that information is not held for all persons who could be involved in senior secondary studies
- students do not choose subjects or schools at random so that within the data set, groups are not random samples from either the participating groups or the underlying complete population
- the factors affecting patterns of participation in subjects and schools are at best imperfectly and incompletely known or understood. Relationships apparent in the data may therefore be partially or wholly misleading.

Given these differences between the data sets involved and those in well-designed research studies, causal or associational inferences, (for example, that this school is more successful than that school) may be completely incorrect.

The principles in this document provide an agreed basis for the release of senior secondary data sets by ACACA agencies. The principles involve recognition that the issues are in essence ethical – we acknowledge our responsibility to our communities for how we select and present data.

Principles

In the release of data, ACACA agencies

- seek to benefit the students, parents, schools and communities with which they work
- comply with any requirements of legislation, applicable law or orders of a court
- act consistently with the protocols under which the data were collected.

When preparing data for public release ACACA agencies have an ethical responsibility to find an appropriate balance of competing considerations:

- their duty to protect students' privacy
- the public interest in the release of useful data
- their duty to take into account how reasonable people are likely to interpret published data.

The protection of privacy requires that data not be released in a form that permits the identification of individual students or their results without their informed consent.

The public interest requires that data be available in meaningful and interpretable forms.

In preparation and presentation of data aggregated at school level, ACACA agencies

- identify the competing balance of validity and reliability in the assessment data collection methods used and declare the extent to which the validity and reliability of the data have been identified
- recognise that reasonable persons will interpret data aggregated at the school level as providing information *about the school,* that is, people expect to draw inferences about the school from school level data
- should therefore only release school level data when the inferences likely to be drawn about the school are reasonable and supported by the data
- recognise that the release of data over time for the "same" groups implies that inferences may be reasonably be drawn about trends and consequently about changes in the performances of organisations involved
- explicitly identify the inferences that such data and its representation in its published form can support and report studies and analyses demonstrating the reasonableness of such inferences
- explicitly identify inferences that are likely to be drawn but that the data do not support and report studies and analyses demonstrating that such inferences are not reasonable
- ensure for any comparisons that the data and its representations suggest may be drawn between groups the uncertainty in the data is quantified and presented similarly and as clearly and prominently as the data themselves

- ensure that any statistical tests of significance explicitly identify non-trivial and relevant null hypotheses
- report the extent of knowledge about patterns of missingness in the data set, including selection biases and their potential impacts on interpretation of the data and analyses of the data.

Strategies for best practice in data release and presentation

Within applicable legislative, policy and existing arrangements, ACACA agencies may

- conduct focus group and related studies with users of the data to understand how the data and its presentations are being interpreted and used
- use a broad range of statistical techniques, including bootstrap studies, for quantifying the uncertainty in estimates derived from aggregated data
- represent the uncertainty in data estimates directly on graphs and tables
- sponsor studies that seek to identify selection biases in the data sets and their likely consequences for the inferences that can be drawn from analyses of the data
- use tables for many numbers
- organise the tables sensibly, arranging rows and columns to display as much structure as possible
- limit the number of significant digits according to the precision of the estimates (for example, where percentages involve a denominator of around 100 it is reasonable to round the reported percentages to 5%)
- use graphs to show complex relationships
- show error bars/confidence intervals on graphs
- avoid pie charts, 3-D histograms and other graphical displays that obscure rather than reveal patterns in data
- find better alternatives to bar charts
- show both cross-sectional and longitudinal patterns with complex trend data
- consider placing into the public domain raw data sets suitably stripped of identifiers and modified with perturbations to protect privacy of individuals and groups while preserving essential features of the relationships in the data set.