

Regarding Statistical Software Support and Priorities at CDC/ATSDR Interim Notes from the Statistical Advisory Group

These notes reflect the intermediate results of SAG's process in developing guidance for statistical software support at CDC/ATSDR.

In late 2011, the Statistical Advisory Group (SAG) was asked to provide a rationale for making decisions about statistical software access. We address a request for guidance on the issue of EpiInfo as a suitable low-cost alternative to SAS, and the larger issue of top tier statistical software access.

Cost Considerations

Our current understanding of the licensing cost of SAS at CDC is that the software is obtained for CDC under a fixed total cost, and that reducing the number of licenses will not alter the total cost – hence, lowering the number of SAS licenses will merely increase the per-license cost. We recommend that EITPO distribute the cost of SAS to the CIOs according to FTEs rather than SAS licenses to eliminate the (false) incentive to restrict SAS licenses.

Regarding EpiInfo

EpiInfo is designed to support the rapid production of data in the field and to allow simple, rapid, basic analysis of this data. From the EpiInfo website (<http://wwwn.cdc.gov/epiinfo/>):

Physicians, nurses, epidemiologists, and other public health workers lacking a background in information technology often have a need for simple tools that allow the rapid creation of data collection instruments and data analysis, visualization, and reporting using epidemiologic methods. Epi Info™, a suite of lightweight software tools, delivers core ad-hoc epidemiologic functionality without the complexity or expense of large, enterprise applications.

EpiInfo serves as a powerful set of tools including data management and analytic support, designed to permit the rapid acquisition of data in public health field applications and to permit rapid, ad hoc analysis of such data. EpiInfo does not serve as an alternative for statistical software such as SAS or Stata.

Regarding a Standard Statistical Software Priority at CDC

Regarding SAS

SAS is the standard software for professional statistical analysis at CDC/ATSDR. SAS provides a full spectrum of support, from data management and manipulation to basic descriptive analysis to complex inference and modeling.

SAS provides a stable, common basis for statistical practice and project management, minimizing training issues and maximizing project continuity. The vast majority of statistical work at

CDC/ATSDR can be conducted entirely within the SAS platform. Common proficiency in a standard software package allows the focus to be on methodology, not platform details.

The use of alternative software for statistical analysis by non-statisticians may be driven by issues of ease or simplicity of use, as well as potential savings, but bear additional costs and consequences:

The support of alternative software will require additional training and support requirements.

The use of multiple software packages will complicate the conduct of statistical work.

While the early work in a project may be well served by an alternative software package, it is likely that substantial projects will grow in complexity, requiring the use of a more complex software package. Using SAS ensures the continuity of project analyses by supporting project work through all levels of complexity.

On the Use of Other Software

Several other statistical software packages are permitted at CDC – there are two broad types of statistical software: comprehensive and application-specific. Comprehensive statistical software packages (e.g., SPSS, Stata, R) support a full suite of statistical routines and functionality, but emphasize different end user needs. For example, SPSS is commonly used in Social Science, Stata supports statistical models and routines frequently used in Economics, and R is preferred by the academic community as it is public domain and allows new statistical methods to be coded and shared quickly without the onerous testing and documentation requirements of commercial statistical software. For a number of statistical applications, there are specialized software packages: SUDAAN (for complex survey work) and StatXact/LogXact (for small sample or exact applications) – these merit support as needed – sample survey and small sample analyses are common special problems in public health applications.

SAG Recommendations

SAG supports a standard statistical software suite for CDC/ATSDR, ensuring a common basis for data management, coding skillsets and project management.

SAG supports SAS as this standard software suite.

SAG supports the allocation of specialized software for specific applications, such as SUDAAN for survey analysis and StatXact/LogXact for exact/small sample analysis as needed.

SAG does not support the use of EpiInfo as a replacement for SAS.

SAG believes that a change in standard statistical software suite for CDC/ATSDR should be discussed at the agency level, not division by division.