



Quick Guide to Good Programming Practice

Getting started:

- Create an initialisation program such as **autoexec.sas**
 - Define **all libnames** and standard options, including **NOFMterr** and **MSGLEVEL=I**
 - Define global and study level macros and macro variables
 - Include a **format program** such as **Studyfmt.sas** where all formats are created
- If the initialisation program is not autoexec.sas, then include the program created at the top of every program
- No libnames or global macro variables should be defined in individual programs
- Create **Abbreviations** for program header, comment and section boxes for consistency

Programming style:

- Use a **standard program header**
- **Keep in mind** while programming: your program will need to be changed, **can you follow it after one year?**
- Set up program into INPUT, PROCESS and OUTPUT sections
- Read in all external datasets **ONLY ONCE**, and do this at the top of the program.
- Only write **ONE** SAS statement per line.
- Use indentation. **Indent 2 characters** (or one tab), but remain consistent throughout the program.
- Use “**KEEP=**” option and specify variable names both when reading in and creating external datasets.
- Use **different meaningful names** for each dataset and variable, avoiding SAS keywords, options and function names.
- Comment programs in the form *** Comments ***; and avoid **/* comments */**.
 - Insert comments in boxes defined in an **abbreviation**. Place comment boxes **above the DATA steps and PROC steps**.
 - Use a different type of box to split the program into sections, so the program is easy to read and follow.
 - **Explain algorithm, logic and reasons** why things are being done in the code below the comment box.

Check the LOG:

- Check the LOG for **ERROR, WARNING, uninitialized** and **repeats of BY values**. The “Repeats of BY values” only appears as a NOTE, but it should be treated as an ERROR and the code must be changed.
- **Variables overwritten** during merge should be corrected so that no variables are overwritten.
- If the LOG contains **Cartesian product** after a PROC SQL statement and a many to many merge was not planned, then check the dataset produced very carefully. It may be incorrect and have more observations than planned.