

Skills for SAS[®] programmers in Epidemiology

Philip R Holland
Holland Numerics Ltd



Skills for SAS[®] programmers in Epidemiology

Agenda

- Size really is important!
- Working with longitudinal databases
- Need to clean the data
- Choosing a suitable database
- Summary



Skills for SAS[®] programmers in Epidemiology

Size really is important!

- Tables measured in Gigabytes, not Megabytes
- WORK disk space usage needs to be checked
- Only keep the columns you need and do this first
- Only sort data once you have your subset
- Don't expect any processing to finish in minutes



Skills for SAS[®] programmers in Epidemiology

Need to clean the data

- Manually selected data causes common mismatches:
 - gender could be male, female, or unknown
 - medical events and prescriptions before practice registration date, or after the date of death
 - prostate cancer in females
 - hysterectomy in males
- Most effort spent in specifying codes to identify medical conditions, or prescription drugs, of interest



Skills for SAS[®] programmers in Epidemiology

Choosing a suitable database - 2

- Databases may not contain all types of data:
 - outpatient and hospitalisation records
 - lab data
 - smoking data
 - repeat prescriptions
 - height and weight data
 - death information
 - links between family members
 - links between medical events and prescriptions



Skills for SAS[®] programmers in Epidemiology

Choosing a suitable database - 3

- Some databases have better representation of certain medical conditions, due to their patient selection:
 - Lifelink has better coverage of Type 2 diabetes than IHCIS due to the higher coverage of elderly patients.
 - US databases include more data on certain drugs than UK databases due to very different prescribing patterns in UK.



Skills for SAS[®] programmers in Epidemiology

Summary

- SAS programmers may need to be able to cope with the following situations:
 - processing massive amounts of data
 - cleaning input data, so that it is consistent, which is not often done by Data Management
 - developing coding list for medical events and prescriptions, again not done by Data Management
 - selecting an appropriate database for each protocol
 - adapting SAS code to cope with differing database data content and structures



Skills for SAS[®] programmers in Epidemiology

Some thoughts for managers...

- Epidemiology data processing is like:
 - detecting fraud in credit card transactions
 - selecting customers for direct marketing
 - analysing web logs
- Epidemiology data processing is not like:
 - clinical trials
- Who should, at least, be interviewed for Epidemiology data processing jobs?



Skills for SAS[®] programmers in Epidemiology

Contact details

Philip R Holland, SAS Consultant

Holland Numerics Ltd

94 Green Drift, Royston, Herts SG8 5BT, UK

tel: +44-7714-279085

fax: +44-1763-244497

email: phil@hollandnumerics.com

web: www.hollandnumerics.com/SASPAPER.HTM

