Associate Editor: Nicola Hills
Short articles (around 600 words), news, tips, quotes, cartoons and other contributions of particular relevance to students are most welcome. Send to: Nicola Hills, c/o the Society's Leicester office. E-mail: Nicola_Hills@hotmail.com

Your data – First things first

BY PETRA BOYNTON

ABOUT a week into my first postdoctoral job I had a nerve-wracking encounter with my manager. We were running a study on intermediate care for older people and planned to collect a large amount of data from questionnaires and patient case notes, and the conversation turned to data management. 'You know about data entry?' my manager asked. 'Sure,' I grinned, happy to be making such a good impression. 'And data cleaning?' 'Yeah,' I nodded, somewhat less confidently. 'How about accounting for people who refuse to participate or aren't suitable for our study? What about missing data?' she added. 'Mmm,' I answered, before scuttling back to my desk.

Although I'd done a PhD, and collected and organised qualitative and quantitative data, I'd only been taught how to analyse it. Not manage it. Data entry was fine, but cleaning it? I'd no idea what that meant, aside from reading through the line of data after I'd entered it and assuming it looked OK. Nobody had taught me how to cope with missing information, or large datasets. And since I had used mainly undergraduates in my PhD, I'd no idea about people being unsuitable for research – students were there to research. As for refusing to take part – who ever says 'no' when you have to complete research for course credits? I was to learn fairly rapidly that non-student samples are slightly more difficult to persuade. I also discovered that no matter how much data you have, if it's missing or incorrectly coded it becomes worthless.

It wasn't exactly my fault. I'd been taught statistics and research and assumed I was as competent as my grades suggested. But like many students I'd not been prepared for the large-scale data collection that may accompany a PhD or research assistant post, or even some undergraduate projects. One crucial tip I picked up was to collect and enter data in 'real time'. For example, in a questionnaire study it is advisable to enter data regularly, rather than collect all the questionnaires before beginning the mammoth task of data entry.

I also wasn't prepared for the amount of work required, or for the necessity of maintaining the quality of my data. Most of us know to be careful and accurate when entering data, but data management goes way beyond this. Remember to double-check all data entered, or check a random number of cases to see whether there's a high level of error (if so, you may need to go through the entire database to ensure accuracy... you never know how mistakes might have crept in). For bigger studies, you can use a double data-entry system, where two people enter the same set of data so errors and inconsistencies can be quickly spotted. Some statistics packages can be set up to play a warning tone if you and your colleague enter different numbers.

Routinely running frequencies can highlight missing cases or numbers that don't belong (e.g. if you have an experiment that produces a maximum score of 15, numbers above this are clearly incorrect and need checking). Missing scores you are aware of are fine, you just need to create codes to account for these. However, missing data you can't explain are a problem, and it adds to your workload to find where they are in your database.

Qualitative 'data' are also important to manage. Again, don't leave all the transcribing to the last minute, and where possible get professional help from a transcribing service (although remember to check what you get back – there may be errors in the transcripts you need to alter).

Data management isn't particularly difficult, although it can be unspeakably dull. However, if you have designed a well-thought-out piece of research, and spent time working with participants, it is only right that you spend as much time ensuring what you code or transcribe is correct. I've included a few core tips here, and there will be more in a book of mine soon to be published. In the meantime, the University of Iowa has produced an excellent guide on 'respecting your data' that covers many issues of data management. See *tinyurl.com/36cz9*.

So whether you're doing a small-scale study or an international project, the rules are the same. Keep your data well groomed and the quality of your work will be that much more reliable.

■ *Petra M. Boynton is in the Department of Primary Care and Population Sciences, University College London. E-mail: p.boynton@pcps.ucl.ac.uk. Her book* The Research Companion: A Practical Guide for the Social and Health Sciences *is due to be published later this by Psychology Press.*

THE SMG IS TEN YEARS YOUNG

To celebrate, we are organising the biggest SMG Conference ever! Why should *you* go? Not only is our guest speaker Professor Simon Baron-Cohen (widely known as a leading expert in the study of autism), but we have also organised special reduced rates: the attendance fee for the Student Conference is £20 for the day (Saturday only) for *all* students.

Booking details and more can be found at *www.bps.org.uk/events/ac2004/smg.cfm*