Getting to know project members

There can be numerous people involved in data collection, such as a project manager, database administrator, database manager, chart abstractors, nurses, data entry personnel, biostatisticians, etc.

Regular interim reports

Do not wait until the end of the project to generate summary reports.

Identify problems. Reports should be generated early and often in order to identify problems in the data entry or collection.

Monitoring reports. There are many types of reports that you may need to generate regularly. For example, you may want reports on the number of records from specific hospitals and abstractors, reasons for excluding certain subjects, number of subjects lost to follow-up due to moving out of state or death, etc. This also allows you to identify outliers and nonsensical results in your data.

Report schedule. A schedule should be kept for generating reports.

Other brochures in this series

- Database ownership (1 of 3).
- Guidelines for detecting bad data (3 of 3).

Related brochures

- How Quantitative Health Sciences can satisfy your research needs.
- Sound principles for simple statistics.
- Working with spreadsheets.

QHS Section

Pippa M. Simpson, PhD
Director

Raymond G. Hoffmann, PhD
Associate Director

Shun-Hwa Li, PhD
Senior Biostatistician

Ke Yan, PhD
Senior Biostatistician

Mahua Dasgupta, MS
Biostatistician

Melodee Nugent, MA
Biostatistician

Chris Cronk, ScD
Senior Epidemiologist

JoAnn Gray-Murray, PHD
Qualitative Researcher

Database Support

Kathy Divine, MS
Database Administrator

Haydee Zimmerman, BA
Database Analyst II

Kim Gajewski, BA
Database Analyst II
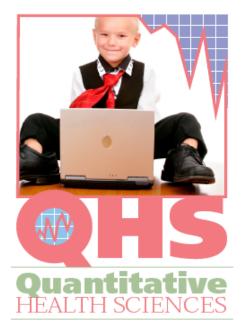
Robert Thielke, PhD
Manager IS II

Children's Research Institute
A member of Children's Hospital and Health System.

MEDICAL COLLEGE OF WISCONSIN

Phone: (414) 955-7675 • Fax: (414) 955-6331
www.chw.org/qhs

Children's Research Institute
A member of Children's Hospital and Health System.

QHS
Quantitative
HEALTH SCIENCES

A good database doesn't mean good data

Avoiding pitfalls that result in bad data

Brochure 2 of 3

Quantitative Health Sciences was established to provide help in the design and analysis of research studies.

Keys to avoiding pitfalls

Backup and recovery (a.k.a. saving yourself from yourself)

Backup procedures should be identified to minimize data loss in the event of hardware and software failures and accidental deletion.

- Backups should be done at regularly scheduled intervals.
- Bulk media backups should be stored off-site (a log may be kept of these locations).
- External hard disks easily can be set up to back up your data while you are away from your computer.

Computer security

No system is completely secure, and precautions should be taken.

Computer viruses. The most current virus detection software should be regularly installed.

Intentional harm. You may feel that you do not have disgruntled employees who may inflict direct harm to your project, but indirect malicious intent can occur. A plan to recover data lost in the event of a disaster should be developed and tested.

Internal access. Group and individual security privileges should be established and maintained for all who participate. Access only should be granted to those actually working with the data. Passwords should expire periodically. HIPAA requirements must be met.

External access. Carefully determine if there is a need for external access and if any harm would result from it.

Data integrity and transaction processing

All data and changes to data should go through a predetermined quality control procedure. It also may be advised to perform manual random comparisons of forms or charts with entered data. You also may use your database to check field ranges, consistency with other records in the data and uniqueness of record identifiers. Other sources of possible error are the original recording, abstraction, data entry and programming.

Abstraction error rates. It is useful to rekey data to get an idea of data entry error rate. A very important error rate to consider is determined by comparing database records to the original chart information. You should generate a random sample of records to be re-abstracted and compared to the first abstraction to get a better knowledge of the abstraction error rate.

Audit trails. Track changes made to the database. This can either be a paper or an electronic trail.

Data queries. Data queries can identify questionable data. If the paper forms indicate agreement with the database, data may have to be checked by re-abstraction.

Rekey or double entry verification. Data entry is a likely time for inaccuracies to occur and can be minimized by rekeying or double entering all of the data. This will allow you to identify inconsistencies or determine an estimate of the data entry error rate.

Describing data flow

The planned flow of data should be communicated and discussed with all involved. A broad description should be included in the study protocol and any details should be described in a database procedure manual. Although each participant may be focusing on different aspects of the project, an overall understanding of data management is the key to effective data management. This also helps determine sources of error, inefficiency and biases.

Documentation

With many people and data involved, there can easily be confusion. *Good documentation helps everyone stay informed.*

With many people and lots of data involved, there easily can be confusion. Good documentation helps everyone stay informed.

Data dictionary or codebook. List each variable with descriptions, formats, allowable ranges, etc.

Procedures manuals. Elaborate on particular processes that may have been briefly described in the study protocol (see below).

Study protocol. Globally describe the project.

Encourage feedback. Obtain feedback from all project members about source of error, inefficiency and bias. Ask for feedback from the beginning and throughout the project.