INFORMATION OVERLOAD - NEW TECHNOLOGIES, CAN WE STORE THE DATA?

Veronica Wiley[1], Kevin Carpenter[2], Bruce Bennetts[3], Bridget Wilcken[1,2]

[1]The NSW Newborn Screening Program; [2]The NSW Biochemical Genetics Service; [3]Molecular Genetics, The Children's Hospital at Westmead, Sydney, NSW, Australia

Abstract. Computerization within newborn screening programs is a developing issue. To date two basic approaches to data storage have been used: (1) a storage system for babies diagnosed with a disorder, (2) a comprehensive system with long-term details for all patient samples, tests performed, test results and interpretations. It usually provides efficient real-time reports for various clinical and quality control requirements and easy access to an inborn errors registry. Within the last decade there have been two new technologies adapted to routine use in newborn screening laboratories: (1) tandem mass spectrometry for selected amino acids and acyl carnitine, and (2) DNA mutational analysis of PCR products. Both technologies present data storage challenges. Both are capable of providing large files of information for a sample. Consideration must be given to how these data are stored, whether all results including a graphical representation or DNA sequence data are kept or whether only final results for specific analytes are stored. Many new analytical technologies can only be incorporated into routine programs as a result of advances in hardware and software allowing better access to, and storage of, data.

INTRODUCTION

Computerization within newborn screening programs is a developing issue. To date, two basic approaches to data storage have been used either as a system for storage of data associated only with babies diagnosed with a disorder or as a comprehensive system. A comprehensive system has long-term storage of details of the patient, sample, tests performed, test results and test interpretations for all babies tested. It usually provides efficient real-time reports, meets various clinical and quality control requirements, and provides easy access to an inborn errors registry.

Within the fields associated with genetic research there are new techniques appearing almost daily. Many of these techniques will have a short life as they fail to live up to expectations and will not convert from a research-based method to routine clinical use. The problem may be that the method required is expensive, the equipment is complicated, or that proprietary reagents are used as consumables. Or, the research method could not easily be converted to a reliable method for clinical situations, ie did not provide sufficient specificity, sensitivity or throughput. Within the last decade there have been two new technologies adapted to routine use in newborn screening laboratories. One technology is tandem mass spectrometry (MS/MS) for selected amino acids and acylcarnitines, and the other is DNA mutational analysis of PCR products.

There are many techniques available for separation and detection of the target sequence of DNA. These include size exclusion polyacrylamide gel electrophoresis with colorimetric stain, allele specific oligonucleotide assay with chemiluminescence or radiographic visualisation, capillary electrophoresis with UV or laser induced fluorescence detection, denaturing high performance liquid chromatography (DHPLC), and more recently, use of mass spectrometry.

Both MS/MS and DNA analysis present data storage challenges. Both are capable of providing large amounts of information about a sample. Consideration must be given as to whether and how these data are stored. This includes whether all results (including graphical representations and DNA sequence data) or final results for specific analytes are stored.

Many new analytical technologies can only be incorporated into routine screening programs as a result of rapidly expanding computer technology. It

is the continuing advances in both hardware and software that allow for better access to, and storage of, data. For programs contemplating incorporation of new technologies, or those expanding from pilot screening to a more expanded population, there are many considerations associated with organising an adequate data storage system that must be considered.

COMPONENTS OF A COMPREHENSIVE COMPUTER SYSTEM

Overview

In order to maximize storage capacity most comprehensive computer systems use relational databases. These databases include structured tables to store data, with each table cross-referenced to the others using a unique identifier. Using indexed fields and applying referential integrity rules that check input in real time improve performance.

Security and information privacy are essential aspects of any computer system. There should be a means for identification of approved users with further authentication (eg passwords). The system should be designed so that data can be retained in a live environment for an extended period. Historic, comprehensive data can significantly contribute to evaluation of the overall screening program. For example in NSW, Australia, a minimum of 10 years of data are available on-line.

A comprehensive data system should include:

- Entry of patient and sample details.
- Ability to match current data with previous data on the same patient (using a unique patient identification number).
- Worklists for analytical procedures that includes repeat samples.
- Result entry including interpretation – manual entry and/or on-line data capture from various instruments.
- Cumulative patient reports and various other user defined statistical and epidemiological reports.
- Quality control and calibrator details.
- Costing information.
- Efficient real-time queries [both structured (query by example) or informal (using structured query language].
- User definable tables with on-line query.
- Disorder registries with additional information as required.

Maintenance

Regular database audits should be performed to ensure the integrity of stored data. These can be minimized if referential integrity rules are utilized in the database scheme. If data storage is considered important, backup of the data is essential. Data backup can occur in many ways and consideration should be given to regular off-site storage of all data.

In order to adequately determine performance indicators (such as screening sensitivity, specificity, positive predictive value and costs) and to establish appropriate local action limits for a newborn screening program, it is necessary to have access to data for each analyte and/or disorder included in screening. Refinements of action limits can be made using alternate limits for selected parameters, for example day of sampling or gestational age, to further maximize performance indicators. For developing programs, it is particularly important to decide the level of computer reliance that is appropriate to the program, keeping in mind any requirements set by the local regulatory certification and accreditation body.

Tables

Various background tables are required for storage of information and can include:

1. User Administration – details of each user with encrypted passwords plus assigned security level determining levels of access to various programs
2. Doctor's Record Administration – contact details for routine or urgent reports
3. Hospital Record Administration – contact details for reports
4. Patient Record Administration – information relevant to the baby including name, sex, date of birth, birth weight, gestational age, multiple birth position, birthplace details and mother's name
5. Sample Administration – information relevant to the sample including date collected, place collected, attending physician, date received in the laboratory, a code for adequacy of sample, comments relevant to the baby's condition at the time of sampling (eg type of feeds etc)
6. Test Record Administration – details for reports including reference ranges and test groups
7. Results – result interpretation for all assays, including those not reported
8. Standards and Quality Controls – details of

calibrators and controls including lot/batch number, details for worklists including position and frequency, and acceptability criteria for results

9. Clinical Comments Administration – a table (library) providing coded comments for reports
10. Inborn Errors Administration – additional details such as genetic consultant, diet and medication
11. Cost – a means of determining the costs of each analytical procedure including costs of reagents, equipment, staffing, etc

Sample logging

The primary function of sample logging is initial entry of demographic information, sample information and required test information. Logging also includes adding new samples to the system. A means for modifying sample and test details in a case where there is relevant clinical information supplied later that requires discriminatory tests should exist. In order for cumulative reporting to be managed efficiently, there must be a means for moving sample information to a different patient identification number when identity was not obvious at the time of initial entry. There should also be a facility to enter relevant clinical information about a patient when needed.

Worklists

All samples awaiting a particular analysis should appear on the current worklist. As tests are requested on samples, the sample number should be automatically queued for inclusion on a worklist. Any sample with abnormal results from a previous assay should also be queued for the next worklist. The worklist format should be appropriate to the analytical method (eg for assays in microtiter format, the worklist should be a grid pattern similar to the plates; while for assays where a linear format is appropriate, the worklist should be a list). The list approach is also suitable for analyses where more than one result is obtained from the same technology. Each worklist should include appropriate calibrators and controls in positions defined by the user.

Result entry

Results should be entered either manually or electronically (a data file generated by laboratory equipment). Result files should be a stored in a quarantine area until all appropriate checking has been completed. As results are entered, computer checks should be performed to ensure that results are within preset user defined limits. If results are outside of the recorded range or appropriate clinical comments are added, then follow-up tests should be automatically scheduled and appear on the next worklist for that assay.

For some newer technologies, eg tandem mass spectrometry, it is also relevant to consider not only individual analyte results, but results compared to other results on the same sample. As long as the comparative results can be freshly generated from raw data, data storage may not be required.

Enquiries

Enquiries can be semi-structured requiring input into predefined fields returning the information in a defined format, or they can be non-structured allowing enquiry of any information stored.

Query by example – An efficient on-line enquiry function allows enquiries using sample number, patient number or any combination of patient's name, date of birth etc.

System Query Language (SQL) – This allows enquiries for any parameter stored within the database and cross-referencing to other parameters via a unique identifier.

Reports

The patient result report format should offer cumulative result reporting and include sufficient identifying details so that no confusion of identity is possible. There should also be an ability to prepare statistical, epidemiological or demographic reports. Examples of reports regularly required include:

1. Patient result report – A cumulative result report usually including the name of the laboratory, the date of report issue, unique patient identifier, date of sample collection, place of sample collection, results, reference intervals, and interpretation
2. Confirmation report – Provides notification to collection points of all samples received over a specified period
3. Sample Receipt Report – A listing of selected details of all samples received
4. Samples received by hospital – A listing of all samples received from a particular source for a defined time period
5. Patients received by hospital – A listing of all babies from whom samples were collected for a defined period

6. Overdue samples – A list of all babies who required a repeat sample collection which has not yet been received in the laboratory
7. Duplicate samples – A list of babies' whose details are the same indicating a possible common identity requiring merging of sample details
8. Performance indicators – An evaluation of program sensitivity, specificity, positive predictive value and costs

Inborn errors registry

Additional information may be required for diagnosed babies. It may be relevant to store additional demographic details for the patient and family as well as for the clinicians involved in treatment. Details of the diagnostic process and subsequent diet and medication can be stored for later retrieval.

CONCLUSIONS

Only with carefully planned and executed computerization can laboratory throughput increase. Whilst data is valuable, storage should be limited so that information that can be easily calculated from new data is not stored. Reports available are only as reliable as the data that are entered and retrieval of data is only possible if it is stored.