

**National Divisions Diabetes Program**



# **Evaluating data quality**

**National Divisions Diabetes Program**  
**Divisions Diabetes & CVD Quality Improvement Project**

*Divisions improving quality of care &  
health outcomes in chronic disease*

November 2003

## 1.0 Purpose of the study

To evaluate the quality of data currently collected by Divisions registers based on the CARDIAB register recall system through a quantitative and qualitative analysis of data quality issues. The study will aim to produce a set of recommendations and proposals that provide a framework for systematically monitoring data quality.

## 1.1 Methods used

Divisions that use CARDIAB as their register recall system and have more than 12 months of data were invited to submit data for the years 2000 to 2002 to the Divisions Diabetes & CVD Quality Improvement Project (DDCQIP). The project adopted a policy framework established under the National Divisions Diabetes Program Data Collation Project in 1999 to ensure that participation was voluntary and that Division data would be used to contribute to quality improvement and feedback mechanisms.

A data cleaning process highlighted potentially invalid data values for Divisions to check. This process identified many potentially invalid data values including high or low HbA1c, high HDL cholesterol, or low Triglyceride levels or missing data. The following template was used:

- (1) Male = Pregnant
- (2) Female < 15 or > 50 = Pregnant
- (3) Type 1 and Rx = Diet or Tablets
- (4) GDM and Rx = Tablets OR GDM and not Pregnant OR GDM Duration >1 year
  - Pregnant male
  - Age <15 or >50 with GDM or pregnant
  - GDM but not pregnant
  - GDM duration > one year
- (5) Year of Diagnosis < Date of Birth OR On Insulin Since Year < Date of Birth
- (6) On Insulin Since Year < Date of Diagnosis
- (7) Age < 2 at Diagnosis
- (8) Age < 20 and on Antihypertensive
- (9) Diastolic BP is > Systolic BP or Systolic <70 or >220 or Diastolic <40 or >130
- (10) Age >14 and Height < 1.3 Metres or > 2.0 Metres
- (11) Age >14 and Weight < 40 Kilograms or > 150 Kilograms OR Age >14 with BMI <15
- (12) Type 1 and On Insulin Since Year >= 3 years after Date of Diagnosis
- (13) HbA1c >20 or <3.5
- (14) Total cholesterol >30 or <2
- (15) HDL cholesterol >2.5 or <0.5
- (16) Triglycerides >50 or <0.5

## Identifying data quality issues

The data cleaning exercise was supplemented by a series of semi-structured discussions with Division officers about the following data quality issues:

- Purpose of Division register/recall system (What is it used for?)
- Data collection process (How are data collected?)
- Types of data error
- Sources of error
- Quality improvement processes

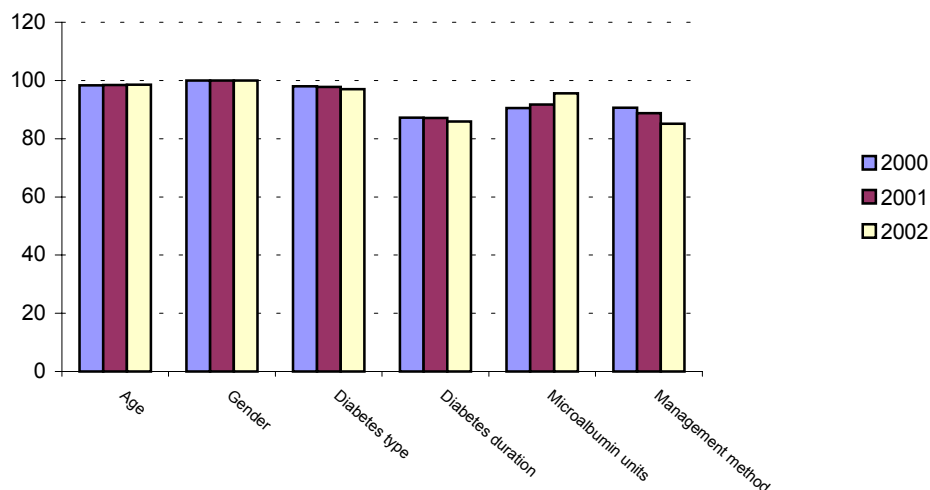
## 1.2 Key findings

### Data completeness

Data completeness is a crucial data quality dimension. The analysis of completion rates for key data variables from Division register data across 16 Divisions showed that age and gender fields showed high completion rates but the variables detailing microalbumin units, management methods and diabetes duration were less than optimal and required improvement.

**Table 1: Completion rates for key variables**

<b>Variable</b>	<b>2000</b>	<b>2001</b>	<b>2002</b>
<b>Age</b>	98.4	98.5	98.6
<b>Gender</b>	100	100	100
<b>Diabetes type</b>	98.1	97.8	97.0
<b>Diabetes duration</b>	87.2	87.1	86.0
<b>Microalbumin units</b>	90.5	91.8	95.6
<b>Management method</b>	90.6	88.8	85.2

**Figure 1: Completion rate for key variables**

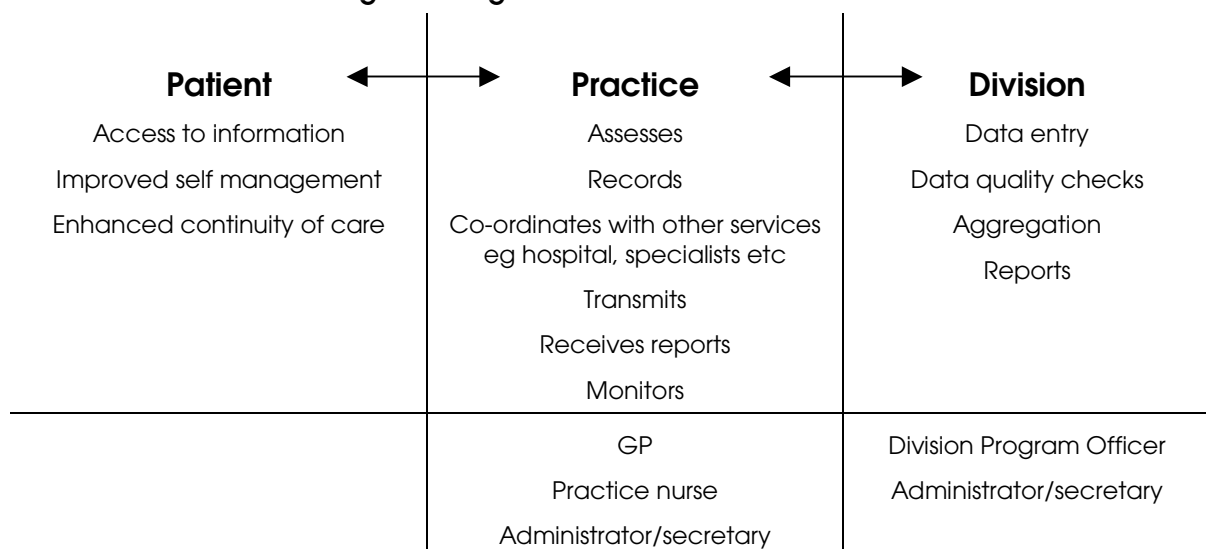
### Dimensions of data quality

Divisions identified a number of key factors affecting data quality:

- Security and confidentiality
- Professional development
- Accountability, communication and ownership

### Data collection process

Divisions revealed that the data collection and collation process could be represented in the following model (see Figure 2) positing the practice as the pivotal part of providing care to the patient, with the Division playing a key supportive role.

**Figure 2: Register recall data collection model**

National Divisions Diabetes Program, Divisions Diabetes & CVD Quality Improvement Project  
 Centre for GP Integration Studies, School of Public Health & Community Medicine  
 University of New South Wales

## Validation and quality assurance

Divisions emphasized the importance of routine data validation to check on completeness of data items and invalid data entries. The data cleaning methodology adopted by the Divisions Diabetes & CVD Quality Improvement Project provides a potential basis for regular data audits. There are a number of types of data error that can be mapped to potential sources.

The following table lists a series of data error types and identifies potential sources for the error.

Type of error	Effect	Potential source		Improving quality of data
		Practice	Division	
Vital data fields missing eg year of birth, year of dx, Type of diabetes, management of diabetes	Incomplete data affecting reliability, and meaning of data.	Data not collected.	Data not entered.	Data audit. Professional development
Potentially invalid Lipids/HbA1c/Hb	Reliability, accuracy and meaning of data	Data incorrectly recorded.	Data incorrectly entered or possibly illegible	Data audit
Type 1 diabetes and Rx diet or tablets	Validity of data	Data incorrectly recorded	Data incorrectly entered	Data validation and professional development.
Potentially invalid data for GDM	Reliability, accuracy and meaning of data	Data not collected or incorrectly recorded.	Data incorrectly entered.	Professional development, data audit and data validation.
Potentially invalid BMI, BP readings.	Reliability, accuracy, and meaning of data	Data not collected or incorrectly recorded.	Data incorrectly entered.	Data audit and validation.

## 1.3 Detailed findings

### Purpose of the Division register/recall system

Divisions listed the following advantages and purposes of their CARDIAB register recall system:

#### Patient management:

- Recall of high risk patients
- Patient management (used to recall patients for investigation and checks)
- Increased community awareness (promotes integration of services)
- Prevention of complications (used as a prompt and guide for GP decision making)
- Promotes patient awareness

#### Planning and service integration:

- Integration with other services
- Program evaluation
- Clinical audits when requested
- Service planning development

#### Reporting mechanisms:

- Reporting to DoHA, management board etc
- Service Incentive Payment support
- Supporting GP management through audits and three month overviews

#### Population health:

- Population health program development
- Targeted GP and patient education
- Research

## Data quality strategies

Divisions identified a number of factors that they described as key to providing the infrastructure and systems for generating quality data and information that users have confidence in.

### Security and confidentiality:

Security and confidentiality are important parts of the proper functioning practice and Division. The rapid development of IT systems has made the collection and dissemination of information easier and less expensive. It has also provided the incentive to collect more data thus increasing the obligation to maintain confidentiality more demanding but still of utmost importance.

The National Divisions Diabetes Program stipulated guiding principles that underpin the collation and aggregation of patient data. These began with the granting of patient consent for data to be collected, the guarantee of patient de-identification, and the right of patients and/or GPs to "opt-off" the system.

Division-based register/recall systems still rely on paper based exchange of information. These methods have associated risks including: loss of confidential information, theft, unauthorised access and misuse of confidential information.

Electronic transfer of records are also liable to risks of loss of confidential information, unauthorised access and abuse of confidential information.

Divisions and practices should adopt a set of guidelines that address the following issues:

- Confidentiality of patient health records
- Safe, secure and well managed filing systems
- IT security (passwords, anti-virus management, access controls)
- Procedures for data backup and disaster recovery

Professional development:

The Division-based register recall process incorporates a range of interactions both within the general practice and with outside bodies including the hospital, other agencies and with the Division. These processes involve the contribution of many people of different professions each with their own training requirements. Divisions identified a number of key areas for professional development:

- Utilisation of evidence based guidelines for the management of diabetes and CVD (Clinicians, Program Officers)
- The importance of register recall systems for implementing guidelines and improving health outcomes (Clinicians, Program Officers and administration staff)
- Knowledge of the purpose, functionality and utilisation of CARDIAB (Clinicians, Program Officers and administration staff)

A number of specific areas involving data quality were identified. These included: understanding data set definitions and standards; software utilisation; awareness of the importance of the data; security and confidentiality and procedures for correcting errors.

Accountability, communication and ownership:

Divisions also identified feedback to GPs as a key validation tool. Reports of results (including aspects of data quality) are a means to stimulate quality improvement and promote clinical ownership of data.

Feedback should be timely, according to agreed deadlines and incorporate the dissemination of appropriate information to all relevant parties. There should also be established procedures for answering queries about data produced for CARDIAB.

## Dimensions of data quality

In order to serve its purpose as a tool in the management of health care, there needs to be the utmost confidence in the quality of data derived from patient encounters. Data quality determines the usefulness of the data – without high quality data, the utility of the information produced is suspect, regardless of the sophistication of the IT system.

The CARDIAB data evaluation study has highlighted the following key attributes of data quality:

- Validity: the data are precise and accurate and are what they purport to be.
- Reliability: the data are recorded in a consistent way using standardised definitions.
- Completeness: all the necessary data are supplied.
- Legibility: the data should be decipherable and legible.
- Timeliness: data are recorded as near as possible to the time of event and made available to those who require the information without undue delay. Quality will be improved if data entry is only carried out once.
- Meaningfulness: data are useful and relevant to the user.
- Accessibility: data are available to those who require them, when they need them.

## 1.4 Implications

### **Consumers:**

Patient control of information is critical to ensuring safety and security. High quality information is vital to safe and effective patient care. Register recall systems that maintain a high level of accuracy can improve patient management services and play a major part in the prevention of complications. Patient education about the role and importance of patient information and data quality is also needed.

Patient and general practitioner security are crucial parts of the proper functioning of register recall systems.

### **General practitioners:**

High quality information is important to maintain the trust of general practitioners in recorded information. It is also vital for achieving safe and effective patient care. Register recall systems can play an important patient management role. They can also serve as tools for clinical audit and evaluation.

### **Divisions:**

High quality information will ensure that the conclusions drawn from data are valid and robust.

Research and population health activities require high quality information.

Divisions can play an important role in providing education to general practitioners and patients about the importance of data quality, including feedback about relevant data quality issues eg data completeness.

### **Policy makers:**

High quality information facilitates understanding of best clinical practice and guidelines and ensures that resources are properly targeted.

Divisions should be supported to provide GP education and patient education about data quality. Division capacity may be an issue.