

# APPENDIX

## APPENDIX 1: DATA MANAGEMENT

### Introduction

Data integrity of a register begins from the data source, data collection tools, data verification and data entry process. Registry data is never as perfect as the clinical trial data. Caution should be used when interpreting the results.

### Data source

The initial phase of the data collected in the Register covered all Renal Replacement Therapy (RRT) patients in the Ministry of Health program since its inception in the early 1970s. The Register subsequently received the data from other sectors of RRT providers like the private, non-government organization (NGO), armed forces and the universities.

The Register continues to actively ascertain new RRT centres in the country. The mechanism of ascertainment is through feedback from the dialysis related company, current Source Data Provider (SDP) and public propagandas. This will gradually and eventually result in a complete RRT centre database. The identified RRT centre is invited to participate in data collection.

Participation in the National Renal Registry which was entirely voluntary prior to 2006 is now made compulsory by the Private Health Care Facilities and Services Act 1996 and its Regulations 2006 which was implemented in 1<sup>st</sup> May 2006. This however only applies to private and NGO centres and data submission from centres managed by the Ministry of Health, Defence or the Universities is still voluntary. RRT centres which have expressed interest in participating will be recruited as SDP.

In the year 2007, there were 44 new haemodialysis centres participated in NRR. Thus, this is an average of 3.6 new centres per month. Within the same year 3 centres had ceased operation. The number of RRT centres is shown in the table below. The participating rate for government centres was 100%. The overall data submission rate has improved from 80.9% in 2006 to 88.8% and with 100% data submission for PD. We hope to see a better participation in the annual treatment return for the coming years.

	At December 2007 Known centres (N)	Agreed to Participate (N)	Submitting data in 2007 (N)	Submitting annual returns (N)	Submitted data (%)
Haemodialysis	453	450	413	357	91.8
Peritoneal Dialysis	32	32	32	30	100
Transplant	70	70	45	45	64.2
All modality	555	552	490	432	88.8

**Data collection**

The data collection tools are designed to mimic the data capture format in the patient case notes to facilitate the data transcription and minimise transcription error. All the SDPs are provided with instructions on data collection and submission to the Register.

The Register collects the RRT patients' demographic details, clinical data, dialysis treatment data, transplant data, peritonitis data and outcome data. The Register holds individual patient's identifiable data that allow complete follow-up despite patient transfers from one centre to another or change of modality which are especially common among the RRT patients. These patients are monitored and tracked through from the time they were registered and commenced their RRT treatment till their death. For those patients who were lost to follow-up, the Register will verify their final outcome with the National Vital Registration System. Patient Profiles are submitted to the Register throughout the year. The identity of patients in the database is not released publicly or in the registry reports.

Centre-specific reports are generated and forwarded to SDP on a quarterly basis. This has generated increased feedback from SDP and improved the patient ascertainment rate and the accuracy of the data transmitted to the Register.

At the end of each year, centres submit their patients' information related to dialysis and drug treatment, clinical and laboratory measurements for the year. Work related rehabilitation and Quality of life Assessment was performed for all patients during the last clinic follow-up.

The Register also conducts an annual centre survey on the staffing and facility profile. The survey questionnaire provides summary information about the number of patients on various treatments. This acts as the basis to calculate the patient ascertainment rate.

**Database System**

The Register initial database was created in DBASE IV in a single computer environment. It was then upgraded to Microsoft Access as a client server application. Currently the NRR data system is a Pentium Xeon 2.4 with dual processors, with a total of 1GB RAM memory and 72GB of RAID-5 (Redundant Array of Independent Disks, level 5). In view of capacity ability, performance and security issues of Microsoft Access, it was subsequently migrated to SQL Server 2000 in the year 2004.

**Data management personnel**

The data management personnel in the Register office are trained base on the standard operating procedures (SOP). The data entry process is also designed to enhance data quality. Quality assurance procedures are in place at all stages to ensure the quality of data.

**Visual review, Data entry and de-duplication verification, Data Editing**

On receiving the case report form (CRF) submitted by SDP, visual review is performed to check for obvious error or missing data in the compulsory fields. Data entry will not be performed if a critical variable on the CRF is missing or ambiguous. The CRF is returned to the SDP for verification.

After passing the duplicate check, the data is then entered and coded where required. Edit checks are performed against pre-specified validation rules to detect missing values, out of range values or inconsistent values. Any data discrepancy found is verified against the source CRF and resolved within the Register office where possible. Otherwise the specific data query report will be generated and forwarded to the SDP to clarify and resolve the data discrepancy.

### **Data coding, data cleaning / data analysis**

Most of the data fields have auto data coding. Those data in text fields will be manually coded by the Register manager. A final edit check run is performed to ensure that data is clean. All queries are resolved before dataset is locked and exported to the statistician for analysis

### **Limitation:**

NRR data submission is still paper base. The majority of the RRT centres do not have electronic patient information system. Computer literacy among staff is still low.

The data submission to the Register is still mainly on voluntary basis using the standard data collection tools. Some SDP choose not to participate in data collection on the patient treatment data for various reasons. We sincerely hope with the enforcement of the Private Health Care Facilities and Services Act 1996 and its Regulations 2006 which was implemented in 1<sup>st</sup> May 2006, participation rate from private and NGO centres shall improve in the coming years.

### **Data release and publication policy**

One of the primary objectives of the Registry is to make data available to the renal community. There are published data in the registry's annual report in the website: <http://www.msn.org.my/nrr>. This report is copyrighted. However it may be freely reproduced without the permission of the National Renal Registry. Acknowledgment would be appreciated. Suggested citation is: YN Lim, TO Lim (Eds). Fourteenth Report of the Malaysian Dialysis and Transplant Registry 2006. Kuala Lumpur 2007

A distinction is made between use of NRR results (as presented in NRR published report) and use of NRR data in a publication. The former is ordinary citation of published work. NRR, of course encourages such citation whether in the form of presentation or other write-ups. The latter constitutes original research publication. NRR position is as follows:

- The NRR does not envisage independent individual publication based entirely on NRR published results, without further analyses or additional data collection.
- NRR however agrees that investigator shall have the right to publish any information or material arising in part out of NRR work. In other words, there must be additional original contribution by the investigator in the work intended for publication.
- NRR encourages the use of its data for research purpose. Any proposed publication or presentation (e.g. manuscript, abstract or poster) for submission to journal or scientific meeting that is based in part or entirely on NRR data should be sent to the NRR prior to submission. NRR will undertake to comment on such documents within 4 weeks. Acknowledgement of the source of the data would also be appreciated.
- Any formal publication of a research based in part or entirely on NRR data in which the input of NRR exceeded that of conventional data management and provision will be considered as a joint publication by investigator and the appropriate NRR personnel.

Participating centre is now able to down load own centre's data from the secured web-site from link from [www.msn.org.my/nrr](http://www.msn.org.my/nrr). Any party who wish to request data for a specific purpose that requires computer-run should make such requests in writing (by e-mail, fax, or classic mail) accompanied by a Data Release Application Form and signed Data Release Agreement Form. Such request will require approval by the Advisory Board before the data can be released.

Distribution of report

The Malaysian Society of Nephrology has made a grant towards the cost of running the registry and the report printing to allow distribution to all members of the association and the source data producers. The report will also be distributed to relevant Health Authorities and international registries.

Further copies of the report can be made available with donation of RM60.00 to defray the cost of printing. The full report is also available in the registry web site [www.msn.org.my/nrr](http://www.msn.org.my/nrr).

## APPENDIX II: ANALYSIS SETS, STATISTICAL METHODS AND DEFINITIONS

### ANALYSIS SETS

This refers to the sets of cases whose data are to be included in the analysis.

Six analysis sets were defined:

#### *1. Dialysis patients notification between 1998 and 2007*

This analysis set consists of patients commencing dialysis between 1998 and 2007. This analysis set was used for the analysis in Chapter 1, 2 and 4.

This analysis set consists of patients with age commencing dialysis less than 20 years old between 1998 and 2007. This analysis set was used for the analysis in Chapter 6.

Since 1993, the MDTR collected annual returns on all dialysis patients to collect data on dialysis and drug treatment, clinical and laboratory measurements. All available data were used to describe the trends in these characteristics. This analysis set was used for the analysis in Chapters 7 to 13.

#### *2. Rehabilitation outcomes*

Analysis is confined to the relevant population. Hence we exclude the following groups.

- Age less than or equal to 21 years
- Age more than or equal to 55 years
- Homemaker
- Full time student
- Retired

This analysis set was used for the analysis in Chapter 5.

#### *3. Centre Survey data*

Section 2.2 in the report was based on annual centre survey data between 1998 to 2007 rather than individual patient data reported to the Registry.

#### *4. Peritonitis data*

Analysis was confined to CAPD patients who were on peritoneal dialysis from 31<sup>st</sup> Dec 1999. This analysis set was used for the analysis in Section 13.4.

#### *5. Economics of Dialysis data*

This analysis used data from on dialysis provision were from the Malaysian Dialysis and Transplant Registry (1980-2005) and international renal provision data from the Annual Data Report of the US Renal Data Service (2005). Published population and economic data was obtained the Department of Statistics, Malaysia Plan reports (1997-2004), World Economic Outlook Database of the International Monetary Fund (1980-2005), World Development Indicators and HNP Stats from the World Bank (1980-2005).

## STATISTICAL METHODS

### Population treatment rates (new treatment or prevalence rates)

Treatment rate is calculated by the ratio of the count of number of new patients or prevalent patients in a given year to the mid-year population of Malaysia in that year, and expressed in per million-population. Results on distribution of treatment rates by state are also expressed in per million-population since states obviously vary in their population sizes.

### Death rate calculation

Annual death rates were calculated by dividing the number of deaths in a year by the estimated mid-year patient population.

### Odds ratio

The odds ratio of an event is the probability of having the event divided by the probability of not having it. The odds ratio is used for comparing the odds of 2 groups. If the odds in group one is 1 and group two is 2, then the odds ratio is 1/2. Thus the odds ratio expresses the relative probability that an event will occur when 2 groups are compared.

With multiple factors, logistic regression model was used to estimate the independent effect of each factor, expressed as odds ratio, on the event of interest.

### Risk ratio

The relative measure of the difference in risk between the exposed and unexposed populations in a cohort study. The relative risk is defined as the rate of disease among the exposed divided by the rate of the disease among the unexposed. A relative risk of 2 means that the exposed group has twice the disease risk as the unexposed group.

### Survival analysis

The unadjusted survival probabilities were calculated using the Kaplan-Meier method, in which the probability of surviving more than a given time can be estimated for members of a cohort of patients without accounting for the characteristics of the members of that cohort.

In order to estimate the difference in survival of different subgroups of patients within the cohort, a stratified proportional hazards model (Cox) was used where appropriate. The results from Cox model are interpreted using a hazard ratio. Adjusted survival probabilities are with age, gender, primary diagnosis and time on RRT used as adjusting risk factors. For diabetics compared with non-diabetics, for example, the hazard ratio is the ratio of the estimated hazards for diabetics relative to non-diabetics, where the hazard is the risk of dying at time  $t$  given that the individual has survival until this time. The underlying assumption of a proportional hazards model is that the ratio remains constant throughout the period under consideration.

Technique failure is defined as occurrence of death or transfer to another modality of dialysis. Similarly, graft failure is defined as occurrence of death or returned to dialysis.

### Analysis of trend of intermediate results

For summarizing intermediate results like continuous laboratory data, we have calculated summary statistics like mean, standard deviation, median, lower quartile, upper quartile and the cumulative frequency distribution graph is plotted over year. Cumulative distribution plot shows a listing of the sample values of a variable on the X axis and the proportion of the observations less than or greater than each value on the Y axis. An accompanying table gives the Median (50% of values are above or below it), upper quartile (UQ, 25% of values above and 75% below it), lower quartile (LQ, 75% of values above and 25% below it). Other percentiles can be read directly off the cumulative distribution plot. The table also shows percent of observations above or below a target value, or with an interval of values; the target value or interval obviously vary with the type of laboratory data. For example, interval of values for prescribed Kt/V is  $\geq 1.3$  and that for haemoglobin is  $<10$ ,  $10-11$  and  $>11$  g/l. The choice of target value is guided by

published clinical practice guidelines, for example, the DOQI guideline; or otherwise they represent consensus of the local dialysis community.

### **Centre survey data**

In contrast to other results reported in this report, Section 2.2 was based on centre survey data rather than individual patient data reported to the Registry. This is to provide an up-to-date information on patient and centre census in the country and thus overcome the inevitable time lag between processing individual patient data and subsequent reporting of results. The survey was conducted in the month of December 2006. Centre response rate to survey was 100%. Standard error estimates are not reported because no sample was taken. Results on distribution by state are also expressed in per million-population since states obviously vary in their population sizes. State population data are based on 2006 census projection. It is very difficult to estimate the amount of cross boundary patient flow; this source of error is therefore not accounted for in computing states estimates. However, we minimize the bias by combining states (Kedah and Perlis) based on geographical considerations. HD treatment capacity is derived by assuming on average patients underwent 3 HD sessions per week and a centre can maximally operate 2.5 shifts per day. A single HD machine can therefore support 5 patients' treatment. Obviously HD treatment capacity is calculated only for centre HD. The ratio of the number of centre HD capacity to number of centre HD patient is a useful measure of utilization of available capacity.

### **Centre variation**

To compare the variation of the intermediate results between centres, graph describing intermediate results in each centre are presented. The 95% confidence intervals have been calculated using the normal approximation of the Poisson to show the variation of proportion in centres. Lower quartile and upper quartile are instead plotted in comparison of variation in median among centres. In the analysis, centres with less than ten patients were combined in a pooled centre. An accompanying table gives the summary statistics like minimum, 5<sup>th</sup> percentile, lower quartile, median, upper quartile, 95<sup>th</sup> percentile and maximum value among centres over year.

Centres with intermediate results for <10 patients were combined into one composite centre.

### **Peritonitis rate**

The occurrence of peritonitis is expressed as number of episode per patient-month of observation; peritonitis rate in short. Relapse peritonitis is defined as peritonitis caused by the same organism occurring within 6 weeks of diagnosis of previous peritonitis.

### **Funnel plot**

Analysis confined to new dialysis patients from year 2000-2007. The figure is included to assess whether survival probability adjusted to age 60 and diabetes of each centre is likely to be different from the national average. Centres with patients less 10 were excluded from the analysis.