

# **APPENDIX I**

## 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 clinical trial data. Caution should be used when interpreting the results.

### Data source

The initial phase of the data collected in the Malaysian Dialysis and Transplant Registry (MDTR) 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.

MDTR continues to actively ascertain new RRT centres in the country. The mechanism of ascertainment is through feedback from the dialysis related companies, 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.

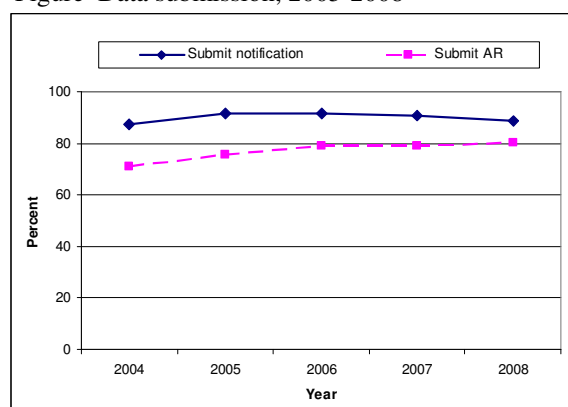
Participation in the MDTR which was entirely voluntary prior to 2006 is now made compulsory by the Private Health Care Facilities and Services Act 1998 and its Regulations 2006 which was implemented on 1<sup>st</sup> May 2006. This however only applies to private and NGO centres and data submission from 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 2008, there were 42 new known haemodialysis centres in Malaysia, i.e. an average of 3.5 new centres per month. One centre ceased operation. The data submission compliance rate for Ministry of Health centres was 100%. The annual treatment data submission has improved among centres with the enforcement of the Act and we hope to see full participation in the coming years. Over all the data submission rate remains good except for renal transplant.

Table Data submission, 2008

	At December 2008 Known centres (N)	Agreed to Participate (N)	Submitting data in 2008 (N)	Submitting annual returns (N)	Submitted data (%)
Haemodialysis	502	489	446	404	91.2
Chronic PD	33	33	32	29	97.0
Transplant	70	70	42	38	60.0
All modality	605	592	520	471	87.8

Figure Data submission, 2005-2008



### **Data collection**

MDTR is a paper base data submission. The case reporting forms are designed to facilitate the data transcription and the information required are readily available in the patient's case note. All the SDPs are provided with instructions on data collection and submission to the Register. The standard data collection forms are colour coded by modality and case report form (CRF) types. The notification forms are submitted periodically or whenever there is an incident. Annual return forms for the assess year should reach the NRR coordinating office not later than January the following year. The CRFs are:

- Patient notification form
- Outcome notification form
- HD annual return form
- PD annual return form
- Transplant annual return form
- Work related rehabilitation and quality of life assessment form – annual assessment

MDTR collects patients' demographic details, clinical data, dialysis treatment data, transplant data, peritonitis data and outcome data. MDTR 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 until their death. For those patients who were lost to follow-up, MDTR 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 transmittal in the registry.

MDTR 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 2.33GHz with dual processors, with a total of 8GB RAM memory and 800GB of RAID-5 (Redundant Array of Independent Disks, level 5). In view of high volume of data accumulated throughout these years, capacity ability, performance and security issues of Microsoft Access, it was subsequently migrated to Microsoft SQL Server 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 than 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 forms. 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 will 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). Sixteenth Report of the Malaysian Dialysis and Transplant Registry 2008. Kuala Lumpur 2009

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 centres are now able to down load their own centre 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 patient notification between 1999 and 2008*

This analysis set consists of patients commencing dialysis between 1999 and 2008. This analysis set was used for the analysis in Chapter 1, 2 and 4.

Patients who were less than 20 years old age at the start of dialysis between 1999 and 2008 were used for the analysis in Chapter 6.

Since 1993, the NRR conducted an annual survey 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. However, in the early years, the data collected from annual survey were relatively incomplete. Hence, for any analysis in relation to these characteristics, we used only data from 1999 onwards when the data were more complete. Remaining missing data in this analysis set was imputed. The raw variables that have been imputed were albumin, calcium, phosphate, hemoglobin, transferrin saturation, cholesterol, ferritin, diastolic blood pressure, BMI and year of birth. This analysis set was used for the analysis in Chapters 7 to 13. However, the generated variable that has been imputed is prescribed Kt/V for HD patients. Prescribed Kt/V which generated are considered the below formula:

$$Kt/V = kdx \times hd\_time \times 60 / (0.58 \times post\ weight \times 1000)$$

where

$$kdx = [1 - \exp(-ex)] \times HD\ flow\ rate \times 500 / [500 - HD\ flow\ rate \times \exp(-ex)]$$

and

$$ex = (500 - HD\ flow\ rate) \times ka / (500 \times HD\ flow\ rate).$$

This variable is considered in Chapter 12.

2. *New Dialysis Patients*

The number of new dialysis patients was based on the first dialysis treatment of the patients. Patients who convert from one dialysis modality to another (from HD to PD or vice versa) are not counted as new patients. If transplant is the 1st RRT treatment and patient's kidney transplant failed and he received dialysis, then for RRT count, the patient will be counted twice. However, if the patient receive transplant in between the dialysis, then the dialysis after transplant will be counted if the transplant last for more than 90 days while if it is last for less than or equal to 90 days, then the dialysis after the transplant will not be counted. This analysis set definition was used in chapters 1,2 and 6.

3. *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 (2007).

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).

International dialysis pricing data was obtained from Harris A. The organization and funding of the treatment of end-stage renal disease in Australia. . Int J Health Care Finance Econ. 7(2-4): 113-132, Hirth RA. The organization and financing of kidney dialysis and transplant care in the United States of America. Int J Health Care Finance Econ. 7(4): 301-318, Nicholson T and Roderick P. International Study of Health Care Organization and Financing of renal services in England and Wales. Int J Health Care Finance Econ. 7(4): 283-299.

International household income data was obtained from Jones F, The effects of taxes and benefits on household income, 2005/06, Office of National Statistics (2007), DeNavas-Walt C, Proctor BD, Hill Lee C. Income, Poverty, and Health Insurance Coverage in the United States: 2005 U.S. Census Bureau (August 2006), Commonwealth of Australia. 6523.0 – Household Income and Income Distribution, Australia, 2005-6. Australian Bureau of Statistics (August 2007)

4. *Rehabilitation outcomes*

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

- i. Age less than or equal to 21 years
- ii. Age more than or equal to 55 years
- iii. Homemaker
- iv. Full time student
- v. Retired

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

5. *Centre Survey data*

Section 2.2 in the report was based on annual centre survey data from 1999 to 2008 rather than individual patient data reported to the Registry.

6. *Peritonitis data*

Analysis was confined to peritoneal dialysis patients from 1999-2008. This analysis set was used for the analysis in Chapter 13.

7. *Renal transplantation data*

This analysis set was confined to patients who had under gone renal transplantation from 1999-2008. This data was obtained from National Transplant Registry (NTR). This analysis set was used for the analysis in Chapter 14.

8. *Diabetes Mellitus*

The patient is considered with to have diabetes mellitus (DM) if the primary cause of ESRD is DM or the comorbid is DM.

## 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.

### ***Primary Renal disease***

Those patients who the primary cause is unknown, pyelonephritis, gouty nephropathy, hypertension or failed transplant, their primary cause will be consider as diabetes mellitus (DM) if their comorbid condition is DM.

*Apply in: Chapter 2, 4 & 14*

### ***Adjusted Mortality of dialysis patients***

Cox propotional hazards model was considered for the mortality of the patients adjusted with demographic and lab variables. This analysis was used in Chapter 4 and 13.

### ***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 by 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 >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. This analysis was used for Chapter 5, 7, 8, 9, 10, 12 & 13

### ***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 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 2008. Centre response rate to survey was almost 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 2007 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 (Selangor and Wilayah Persekutuan Putrajaya, Sabah and Wilayah Persekutuan Labuan) 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. This analysis was used in Chapter 2.

### ***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, 5th percentile, lower quartile, median, upper quartile, 95th percentile and maximum value among centres by year.

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

This analytical method was used in Chapters 7, 8, 9, 10, 11, 12 & 13

### ***Death rate***

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

### ***Incidence rate ratio***

The incidence rate is determined by dividing the number of new cases of a diseases or condition in a specific population over a given period of time by the total population. Therefore incidence rate ratio is the comparison of two groups in terms of incidence rates. Poisson regression model was considered to estimate the independent effect of each factor, expressed as incidence rate ratio. An incidence rate ratio of 3 means that group 2 have the rate 3 times higher than group 1 when group 1 is the reference group.

### ***Odds ratio and variation in odds ratio 2007***

The cohort considered for this analysis was patients who were on dialysis in 2005 and new patients in 2005.

The odds 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 1 is 1 and group 2 is 2, then 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 such as dialysis center, age, sex, modality, albumin, hemoglobin, calcium, cardiovascular and cholesterol, logistic regression model was used to estimate the independent effect of each factor, expressed as odds ratio, on the event of interest and the variation is odds ratio. This method was used in chapter 4.

### ***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 adjusted for age, gender, primary diagnosis and time on RRT. 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.



***Patient survival was considered in two ways:***

- i. Survival censored for change of modality based on the first modality. Duration of change modality or transplant will not be considered.
- ii. Survival not censored for change of modality. Duration survival for patients will be accumulated from the first till last treatment received. The duration of treatment with any change of dialysis modality or with transplant will be considered.

***Survival of incident patients by centre***

***1 year survival***

The cohort consider for this analysis was considered from 1999-2007. Many patients commencing dialysis in 2008 would still not have completed one year.

***5 years survival***

The cohort consider for this analysis was considered from 1999-2003. This is due to those commence from 2004 onwards still not able to have 5 year survivals analysis.

***Funnel plot***

This analysis was confined to new dialysis patients from year 2000-2007. The figure is included to assess whether survival probability adjusted to age 60 and diabetes for each centre is likely to be different from the national average. Centres with patients less 10 will be excluded from the analysis. This plot was used in Chapter 4.

***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.