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 1995 and 2004***
   This analysis set consists of patients commencing dialysis between 1995 and 2004. This analysis set was used for the analysis in Chapter 1, 2 and 3.

2. ***Dialysis patients notification between 1990 and 2004***
   This analysis set consists of patients with age commencing dialysis less than 20 years old between 1990 and 2004. This analysis set was used for the analysis in Chapter 5.

3. ***Dialysis patients between 1997 and 2004***
   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 analyses in relation to these characteristics, we used only data from 1997 onwards when the data were more complete. Remaining missing data in this analysis set was imputed using first available observation carried backward or last observation carried forward. This analysis set was used for the analysis in Chapters 6 to 12.

4. ***Rehabilitation outcomes***
   Analysis is confined to the relevant population. Hence we exclude 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 4.

5. ***Centre Survey data***
   Section 2.2 in the report was based on annual centre survey data between 1999 to 2004 rather than individual patient data reported to the Registry.

6. ***Peritonitis data***
   Analysis was confined to CAPD patients who were on peritoneal dialysis from 31st Dec 1999.
   This analysis set was used for the analysis in Section 12.4.

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 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 O1 and group 2 is O2, then odds ratio is O1/O2. 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.

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 2004. 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 2004 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, 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^{th} percentile, lower quartile, median, upper quarter, 95^{th} 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.