Appendix B: Definition, Statistical Methodology, Analysis Criteria

B:1 Definitions of analysis quarters

Quarter	Dates
Quarter 1	1 January–31 March
Quarter 2	1 April–30 June
Quarter 3	1 July–30 September
Quarter 4	1 October–31 December

The quarterly biochemistry data are extracted from renal unit systems as the last data item stored for that quarter. If the patient treatment modality is haemodialysis, the software will try to select a pre-dialysis value.

B:2 Renal Registry modality definitions

Home haemodialysis

Home haemodialysis patients cease to be classed as such if they need longer than 2 weeks of hospital dialysis when not an inpatient.

Satellite dialysis unit

A renal satellite unit is defined as a haemodialysis facility that is linked to a main renal unit and not autonomous for medical decisions and that provides chronic outpatient maintenance haemodialysis but with no acute or inpatient nephrology beds on site.

Treatment modality at 90 days

This is used by the United States Renal Data System (USRDS) and is the modality that the patient is on at day 90 regardless of any changes from the start. It is a general indicator of initial dialysis but could miss failed CAPD. This would also miss patients intended for home haemodialysis who were not home yet. This modality is calculated by the Registry, which allows the definition to be changed.

Start of established renal failure (ERF)

Established renal failure (also known as end-stage renal failure/end-stage renal disease) is defined as the date of the first dialysis (or of pre-emptive transplant).

If a patient is started as 'acute' renal failure and does not recover, the date of start of renal replacement should be backdated to the start of acute dialysis.

If a patient is started on dialysis and dialysis is temporarily stopped for less than 90 days for any reason (including access failure and awaiting the formation of further access) except the recovery of renal function, the date of start of renal replacement therapy (RRT) remains the date of first dialysis. If the patient has stopped for longer than 90 days, he or she is classed as 'recovered'.

B:3 Analysis criteria

Definition of the take-on population (incidence)

The take-on population in a year included patients who later recovered from ERF after 90 days from the start of treatment. Patients newly transferred into a centre who were already on RRT were **excluded** from the take-on population for that centre. Patients restarting dialysis after a failed transplant were also excluded (unless they started RRT in that current year).

Since patients who restarted RRT after recovering from ERF are included in the take-on population, the following scenario can occur: a patient may start RRT in 2003, recover and then restart RRT in 2003. Such patients are counted twice in the analysis providing they have been receiving RRT for more than 90 days on each occasion.

Patients who started treatment at a centre and then transferred out soon after receiving treatment are counted at the original centre for all analyses of treatment on the 90th day.

Definition of the prevalent population

This is calculated as all patients who are alive on 31 December and includes the incident cohort for that year alive on that date.

Confidence Interval

The 95% confidence intervals have been calculated using the normal approximation of the Poisson.

Death rate calculation

The death rate per 100 patient years was calculated by counting the number of deaths and dividing by the person years exposed. This includes all patients, including those who died within the first 3 months of therapy. The person years at risk were calculated by adding up, for each patient, the number of days at risk (until they died or transferred out) and dividing by 365.

Odds ratio

The odds of dying is the:

$$\frac{\text{(Probability of dying for someone with a phosphate of 1.71}-\text{2.10 mmol/L)}}{\text{(Probability of surviving for someone with a phosphate of 1.71}-\text{2.10 mmol/L)}}$$

The odds ratio is the:

$$\frac{\text{(Odds of dying with a phosphate of 1.71}-\text{2.10 mmol/L)}}{\text{(Odds of dying in the reference group)}}$$

Hazard function

The hazard function is the probability of dying in a short time interval considering survival to that interval.

Hazard ratio

$$\frac{\text{(Probability of dying in the next interval for a phosphate of 1.71}-\text{2.10 mmol/L)}}{\text{(Probability of dying in the next interval for a phosphate in the reference range)}}$$

Z-Scores

The enquiry into the excess of paediatric cardiac deaths at the Bristol Royal Infirmary defined an outlier as lying beyond 3 standard deviations from the mean, using the statistical methodology of Shewhart's control theory. This analysis relies on the centre sizes, and hence their standard deviation, being very similar. Renal units in the UK vary greatly in size, catchment populations varying from 300,000 to over 2 million. There is a consequent variation in the total patient number on RRT so the figure for the standard deviation will vary greatly between centres. The standard deviation for the total RRT population is not an appropriate number as this will be very small. Therefore, the Shewhart methodology cannot be applied. The Registry has used the accepted statistical technique of Z-scores to identify any outliers.

Definition

Z-scores are sometimes called "standard scores". It is a measure of the distance in standard deviations of a sample from the mean.

The Z-score transformation is especially useful when seeking to compare the relative standings of items from distributions with different means and/or different standard deviations. The Z-score for an item indicates how far and in what direction, that item deviates from its distribution's mean, expressed in units of its distribution's standard deviation.

Mathematically:

$$\text{the survival Z-score} = \frac{\text{Survival for centre X} - \text{survival for all centres}}{\text{Standard error for centre X}}$$

The Z-score is therefore an adjustment for the size of the centre and when comparing the different Z-scores for all the centres, they should be normally distributed. The observed Z value compared with the expected Z value (see explanation below) should be on a straight line.

Calculation of the expected Z value

Suppose there is a normally distributed population from which we repeatedly draw random samples of some specific size, say 10. These 10 values from each such random sample are sorted into increasing order, smallest value to largest value. When the sample data is sorted in this way, the individual numbers are called **order statistics**. The smallest value will vary somewhat from one such sample to another, but over the long run, the smallest values should tend to cluster around some average smallest value and produce a **mean** or **expected values of the order statistics**. These data have been compiled into tables so that for every specific total number of ordered samples (eg 38 centres with Registry survival data)

there is an expected Z value for each ordered centre in that list.

Survival analyses of prevalent cohort

These analyses exclude the current year's incident cohort. Note some Renal Registries include these patients in the prevalent survival.

Criteria for analysis by treatment modality in a quarter

The following quarterly entries were included and excluded:

- Patients on haemodialysis with a treatment centre of 'elsewhere' were **removed**. It should be noted that there were some patients on transplant with a treatment centre of 'elsewhere'; these patients were **included.**

- Entries for which the hospital centre was not the primary treatment centre were removed from the analysis of data for that centre.

- Patients who had been on RRT for less than 90 days were removed (by definition of ESRF).

There were, however, a few exceptions to these rules:

1. If a patient's initial entry on the treatment timeline contained a **'transferred in'** code, the patient was assumed to have been on RRT for longer than 90 days since the patient must have started RRT earlier than this elsewhere. Therefore, patients with an initial entry on the treatment timeline with a **'transferred in'** code were included for all quarters. A patient with an initial treatment modality of **'transferred in'** on 1 March 2003 would, for example, be included for the quarter 1/03 even though the number of days on RRT would be calculated as 30 days.

2. For patients who **recovered** renal function for a period of time and then went into ESRF, the length of time on RRT was calculated from the day on which the patient restarted RRT. For example, the number of days on RRT would be calculated from 1 November 2003. The patient would be excluded from the analysis for quarter 4/03 since on 31 December 2003, he or she would have been on RRT for only 90 days. The patient would be included in the analysis from quarter 1/04 onwards.

If recovery was for less than 90 days, the start of renal replacement therapy will be calculated from the date of the first episode and the recovery period will be ignored.

Patients who had **transferred out** or **stopped treatment without recovery of function** before the end of the quarter were excluded.

Criteria for analysis of biochemistry in a quarter

The analysis used information from the quarterly treatment table. In addition to the treatment modality criteria listed above, patients with the following quarterly entries were also excluded:

1. Patients who had **'transferred in'** to the centre in that particular quarter were excluded. If, for example, a patient transferred in on 1 March 2003, the patient was excluded from that biochemistry analysis of the centre transferred to, in that quarter.

2. Patients who had changed treatment modality in that particular quarter were excluded.

Treatment modality on day 90 of starting RRT

This is obtained from the treatment modality of the take-on population after 90 days of being on RRT. For this reason, patients who started treatment between 1 October 2002 and 31 September 2003 were used in this analysis.

The sample used was that defined by the take-on population.

Patients were counted at their take-on hospital centre rather than at their hospital centre on day 90. This is important as some patients had transferred out of their initial hospital centre by day 90.

Patients who died before they reached 90 days were excluded.

One-year survival of the take-on population

The sample used was the same as that defined for the take-on population except for recovered renal function patients who were excluded.

Patients who transferred out of their initial treatment centre were censored on the day they transferred out if there was no further information in the timeline.

Analysis of 1 year survival of prevalent patients

The death rate within the year was calculated separately for the patients established on dialysis and with a functioning transplant on 1 January 2003. As there is an increased death rate in the first 6 months following transplantation, patients were included in the analysis only if they had not received a transplant between 1 July 2002 and 31 December 2002. For the same reason, patients who received a transplant within the year were censored at the time of transplantation.

The sample criteria thus became:

1. Patients who had been receiving RRT for more than 90 days on 1 January 2003.

2. Patients who had a transplant between 1 July 2002 and 31 December 2002 were excluded.

3. Patients who transferred into a Registry centre were excluded if information was not available to confirm that they had not received a transplant between 1 July 2002 and 31 December 2002.

4. The few patients who recovered renal function in 2003 were excluded.

5. Patients who transferred out of a Registry centre to a non-Registry centre were censored at that date.

6. A transplant patient whose transplant failed was censored at the time of restarting dialysis and dialysis patients who received a transplant were censored at the time of transplantation.

7. Patients who died, received a transplant, or transferred out on 1 January 2003 were included and were counted as being at risk for 1 day.

8. Patients who died on the day of the transplant were censored on this day rather than counted as a dialysis death.

Seasonal variation of deaths using circular data statistical technique

In a study with a cyclic time period, such as a year, it is possible to interpret these data as circular data. (Mardia, Statistics of Directional Data, 1972)[1]. Circular analysis has advantages over linear analysis as circular data has no beginning or end. Circular data cannot be ordered for example, December is not 'larger' than January.

An example of the importance of circular data analysis over linear analysis would be looking at the observed angles in a data set, 15° and 354°, the arithmetic mean would be approximately 180°, and this is clearly not the average direction. This is illustrated in the figure B.1.

Hence, why it is better to use circular analysis, the mean direction is 0° and this would be the resultant date of death. Analyses carried out with this methodology need to be specific to circular data.

The patients included were an incident cohort of all those patients who died between 1997 and 2003. Data used are the Townsend score, age and additional data on the daily temperatures from 1997 to 2003 are also included. The response variable of interest is the date of death. Each day in the year represented as an angle on a circle results in a set of circular data.

The seasonal variation in deaths were assessed using standard uniformity tests in the form of the Rayleigh Test. Rayleigh's test is a parametric test which assesses the significance of the mean resultant length which is the measure of strength of the mean direction. Any departure from uniformity and evidence of true seasonal variation would be shown

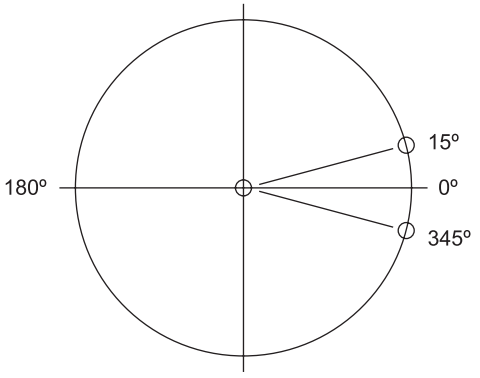

Figure B.1 Analysis of circular data

here. The mean direction gives the 'preferred' day of death for a patient on RRT.

Circular-regression analysis is carried out to assess the affect of different variables on the date of death. The response variable is the date of death and is a circular variable.

Reference

1. Mardia KV. (1972). *Statistics of Directional Data*. Academic Press: London.