Technical Annex

About the Kidney PREM 2019 Data

As in 2018, the Kidney PREM data consisted of 39 questions covering 13 themes plus an overall experience question. Patients responded to each question on a scale from 1-7. All questions had the option of "don't know" and "not applicable", with the exception of question 39 "Your Overall Experience". The themes and related questions with the response scale can be seen in table 1:

Table 1: Themes in the 2019 Kidney PREM, with the response scale.

Section	Theme	Questions	Response scale
1	Access to the Renal Team	Q1-Q3	1 Never – 7 Always
2	Support	Q4-Q6	
3	Communication	Q7-Q11	
4	Patient Information	Q12-Q13	
5	Fluid Intake and Diet	Q14-Q15	
6	Needling	Q16	
7	Tests	Q17-Q19	
8	Sharing Decisions About Your Care	Q20-Q22	
9	Privacy and Dignity	Q23-Q24[a]	
10	Scheduling and Planning	Q25-Q26, Q27[a]	
11	How the Renal Team Treats You	Q28-Q30	
12	Transport	Q31-Q33	
13	The Environment	Q34-Q38	1 Poor – 7 Excellent
14	Your Overall Experience	Q39[b]	1 Worst it can be – 7 Best it can be

Those questions marked (a) referred to filtered questions, where only a subset of patients was required to answer. Question 39 (b) was a question about patients' overall experiences. In addition to the survey data, information about patient characteristics was collected for current treatment (including CKD and transplant, and location if haemodialysis), age category, gender, ethnicity and use of PatientView.

Data Collection Process

Online data collection was over a period of 8 weeks throughout June and July. Collection of paper copy data was advertised over a period of 4 weeks, from 1st to 30th June. Whilst the online survey was closed on 31st July, Kidney PREMs were returned to UKRR up until 31st August, and so may have been completed outside of this window.

Paper Questionnaires

Each centre across England, Wales, Scotland and Northern Ireland was invited to take part in the survey, with some units collecting data from a larger proportion of patients than others. Paper questionnaires were made available in units for patients to complete, with a member of staff inputting their UK Renal Registry renal unit code on the paper questionnaire.

Completed questionnaires were sent to the Renal Registry where they were scanned. UKRR scanned 15,530 paper questionnaires, but of these 220 had 0 questions answered and a further 22 with only 1 question answered (assumed to be a scanning error). A further 37 were found to have been

scanned incorrectly during quality control and were in fact blank. Therefore, 279 questionnaires were removed from the data, resulting in a dataset of 15,251.

Online Questionnaires

The Kidney PREM survey was also made available online via the Renal Registry website (https://www.renalreg.org/projects/prem), with patients able to select their unit from a drop-down list. Space was given to type a unit name if patients were unable to find the correct unit. Communication about the availability of the online survey was not applied consistently across units and some patients may have found the survey online without being told it was available. 1,583 patients submitted an online questionnaire, but of these 365 had answered 0 questions. These surveys were removed from the data, resulting in a dataset of 1,218. In addition to English, the questionnaire was available in Welsh, Gujarati and Urdu, although all questionnaires were completed in English.

The number of total valid responses to the Kidney PREM 2018 was 16,469 (15,251 paper and 1,218 online).

Data Cleansing

All data analysis was done using either Excel or Stata/IC version 15.1. UKRR codes were added to the online data based on the unit selected by the patient from the drop-down list. 163 records had entries in the "Unit Free Text" box either in addition to the unit selected or instead. UKRR checked these records; 135 were subsequently updated. The paper Kidney PREM results were appended to the online Kidney PREM results, with a variable created to identify the origin of each record (paper or online). Patient characteristics variables (treatment, treatment location, age, sex, ethnicity and PatientView use) were encoded to allow for analysis. UKRR codes and hospital/unit names were checked and amended where appropriate for consistency. The patient summary characteristics table was produced based on all patients who provided valid responses and percentages were calculated across each characteristic (age, gender, ethnicity, treatment and haemodialysis location).

Estimation of Scale/Sub-Scale Scores

Sub-scale scores were estimated for each theme, for each patient. This was to give a score for each theme, with equal weight given to every question. If questions were not answered then theme sub-scale scores could still be estimated, so long as there was no more than 1 question missed. For themes containing just one question a score could not be estimated. Rules were applied to questions within themes 6 (Needling) and 10 (Transport) as follows:

1. Only in-hospital and in-satellite HD patient answers were analysed for Q16 (Needling)
2. In-hospital and in-satellite HD patient answers excluded for Q27 (blood tests)
3. Patients where treatment type was missing excluded from analysis for both questions
4. Patients who selected HD without specifying a location assumed to be in-hospital/in-satellite and analysed accordingly since very few home HD patients would have completed the survey.

Theme 10 (Scheduling and Planning) contained only one filtered question out of three, so scores were estimated using the unfiltered questions if applicable.

The overall scale score was estimated excluding question 39 (Your Overall Experience). This left 38 questions, of which five were filtered leaving 33 for inclusion. Scale scores were estimated if there were less than four missed questions from the 33, excluding those where "Don't Know" or "N/A" was selected (approx. 9%).

Sub-scale and scale scores were not estimated for patients with missing centres, affecting 29 patients who had answered at least 4 questions, in line with previous years' analyses.

Sub-scale and scale scores were each calculated using the following algorithm:

- "Don't Know" and "N/A" responses were recoded as missing
- Number of missed responses from each theme and the overall score was calculated (from unfiltered questions) (M)
- Total score for each theme and the overall scale was calculated (R)
- The scale or subscale core was calculated (if number of missed responses ≤1 for each theme or ≤4 for overall score): $= \frac{R}{Q-M}$ where R is the number of questions being evaluated

Mean Scores by Centre

Mean sub-scale and scale scores were calculated across each centre. Scores were only reported when there were seven or more responses per centre. One centre had fewer than seven responses throughout, plus an additional centre had insufficient responses for the Needling and Transport themes.

The means estimated in this way were used for reporting in the following tables in the main body of the report or available separately online:

- Table 1, comparing 2019 to 2018 data (for those centres with seven or more responses);
- Table 4, a summary of the highest and lowest mean scores by centre, with the range in centre scores, for 2019, 2018 and 2017;
- Graphs of mean 2019 theme scores by centre; and
- Graphs showing 2018/19 theme comparisons (available online).

The caterpillar plots in the Kidney PREM report (Mean 2019 theme scores by centre) provide a visual guide to variation between centres across the 13 themes, and for the overall question. Each plot (one per theme) shows the median, lower quartile and upper quartile for all centres as a vertical line. For each theme, the data was sorted in descending order by centre, and the mean value for each centre and 95% confidence intervals are shown. Centres with less than seven responses for any theme were excluded from the graphs.

For the waterfall plots (2018/19 theme comparisons), mean scores across centres were calculated for each theme. 2018 mean scores were also calculated (excluding centres with fewer than 7 responses). The data was sorted in descending order of the 2019 means, and by centre. The plots show the 2018 and 2019 means for each centre and theme, with the overall mean value for 2018 and 2019 as vertical lines.

Question Response Centre/Unit Data

In addition to this report, question level response data was made available to centres via an online portal. Some relabelling or recoding was done to improve the readability of the dataset. Patient answers of any unit or centre with less than 10 responses were removed to preserve the anonymity of individual patients (note the plots exclude centres with 7 or fewer responses).

Data in the portal is grouped by geographical location (Country, region, main unit and site). Data presented in the bar-chart and the table can be separately expanded or contracted to amalgamate sites, centres, regions or countries by using the small (+) and (-) symbols which appear when a user

hovers over the geography title in the chart or table. In addition it is possible to restrict the data to regions, centres or satellites using the filters to the right of the table/chart. Individual questions are selected individually using the panel under the table. It is not possible to select multiple questions simply because of the volume of data which would potentially being displayed within the panels. Data in the portal is presented as either numbers of people who gave each response (one to seven, not applicable, don't know or missing), or as a proportion of total group who gave a numerical response (i.e. excluding the NA, don't know and missing responses). In both cases hovering over the column or the cell displays both the value and proportion in a tool-tip. It is crucial to consider the number of people making a response before making a judgement on whether the proportion who responded in that way is large enough to allow firm conclusions to be drawn.

In addition to the portal data, tables were produced displaying:

- Means of each theme by treatment modality (overall)

For each centre:

- Means and 95% confidence intervals of each theme by treatment modality
- Means of each theme by treatment modality for 2019, 2018 and 2017.

For each item under consideration, data was removed if there were fewer than 7 responses, to limit the potential of patient identification.

<u>Limitations and caveats for interpreting the plots and data tables:</u>

Presenting data to the community to allow for meaningful interpretation is always a challenge. As with last year, additional tables have been provided for each centre providing means and confidence intervals for each of the 13 themes, and for each treatment modality, adding to the information provided by the caterpillar and waterfall plots. Any summary of data (means, intervals) leads to loss of information but increases the ability to make sense of trends across different groups.

The Kidney PREM 2019 data is challenging as the distribution of responses across the response options (1-7) does not follow a "normal" distribution. People tend to use the 5, 6 and 7 response options much more than 4 or less, although a considerable number of people do wish to report a poor experience (referred to as a skewed distribution). A common way to deal with a skewed distribution is to use a median with quartiles to display the distribution of the data. However, if the median is estimated for the Kidney PREM data most questions and themes have a median of 7, and sometimes 6, meaning that the median as a way of communicating the central tendency is not sensitive to variation between questions or themes.

In addition, we are in the happy position of having a very large number of responses from the Kidney PREM. This means that the statistical reasons for reporting the median and quartiles is less important, and the mean and 95% confidence interval provides a robust picture of the responses for most people at the top end of the scale and makes it much easier to compare different groups. By examining the mean and confidence intervals we can have a high level of confidence that the intervals capture the responses of the majority of any group. We can therefore be comfortable that if the interval for a particular group falls below, or above the 25^{th} or below the 75^{th} percentile for the group as a whole, then the majority of people responding group will be within that range, and the group can be considered to fall below or above the relevant percentile.

However, the choice of a mean and 95% confidence interval does mean that the confidence interval gives us less information about the "tail" of responses in the lower parts of the response scale. This should be borne in mind when making interpretations of the data. The width of confidence intervals are sensitive to sample size. A confidence interval is inversely proportional to the square root of the sample size (the confidence interval is produced by dividing by the square root of the sample

size). This means that as the sample size increases the confidence interval gets smaller. If everything else is the same, the confidence interval for a sample size of 15 will be twice as big as a sample size of 60 (e.g. interval/$\sqrt{4}$). This is critical for small sample sizes of less than 30 where the confidence interval is likely be very large. In a very real sense, a large confidence interval indicates uncertainty which is reflected in the idea that a small sample size is not representative of the population. On the other hand, a very small confidence interval may simply reflect a very large sample size and give the false impression of difference where the clinical or psychosocial meaning is less obvious.

Interpreting the number of responses:

It is particularly important that where there are a small number of responses that caution is taken in interpreting what this may mean. It is common for numbers to be translated into percentages, but this may be misleading.

As an example, we may think about testing a new drug. If 3 out of 5 patients respond well to the drug, can we say that the drug works for 60% of patients? We know that in these circumstances there is an element of chance. If we take another group of 5 patients and only 2, or perhaps 4 patients respond to the drug, can we say that actually the drug works for 20%, or 80% of patients? The problem is the uncertainty related to small numbers. Where we have a test of 100 patients, then we can have more confidence in the observed numbers of responses, say 56/100. But even then, the chance of another sample of patients differing by as many as ±10 is considerable. Just looking at the percentage of responses in a particular group, without considering the number of responses may be very misleading.