

FULL METHODS

We constructed a survey to assess the likelihood that nephrologists would recommend patients initiate dialysis. Two versions of the survey were created that were identical except one version included only serum creatinine values (Creatinine version) and the other included both serum creatinine values and eGFR (eGFR version). Potential respondents were randomly allocated to receive either the Creatinine or eGFR version. Respondents were not informed that there were two versions of the survey and they explicitly consented to the use of their responses in research studies. The study was approved by the St. Josephs Healthcare Research Ethics Board in Hamilton, Ontario, Canada and the John F. Wolf, M.D. Human Subjects Committee at the Los Angeles Biomedical Research Institute in Torrance, California, U.S.A.

Development of survey instrument

The survey was designed by four practicing nephrologists and after multiple iterations amongst the designers the survey was piloted by five nephrologists not involved in its design. The pilot assessed the survey for face validity, content validity, clarity, utility, discriminability and item redundancy. After final modifications, the survey was also reviewed by investigators from each target country for clarity.

We constructed a hypothetical scenario in which a patient with chronic kidney disease developed moderate symptoms of uremia¹⁰. For each scenario 4 levels of renal function were provided as both a creatinine concentration and eGFR based on that creatinine using the 4 variable Modification of Diet in Renal Disease equation¹¹. Renal function corresponded to an eGFR of 5 ml/min/1.73 m², 7.5 ml/min/1.73 m², 10 ml/min/1.73 m², and 12.5 ml/min/1.73 m². Respondents were then asked to rate the likelihood they would recommend initiation of dialysis.

A response was recorded on a modified 8-point Likert scale anchored at “Definitely Not”=1 and “Definitely Would”=8. The scenario was then repeated with the patient’s comorbidities adjusted to reflect moderate and severe levels¹². Two versions of the survey were then constructed, one that included serum creatinine only (Creatinine version) and one that included both serum creatinine and its corresponding eGFR (eGFR version) for each question.

To assess the relative importance of eGFR reporting we developed an additional scenario in which the degree of uremic symptoms were modified from moderate to mild (i.e. fatigue only) with an eGFR of 10 ml/min/1.73 m². Respondents were asked the likelihood of recommending dialysis on the same modified Likert scale as the other scenarios.

Survey Sample

Lists of email addresses of potentially eligible nephrologists were assembled by liaising with a professional society in each target country (the Canadian Society of Nephrology in Canada, the National Kidney Foundation in the USA, the Renal Registry in the UK, and the Australian and New Zealand Society of Nephrology for Australia and New Zealand). These membership lists include non-nephrologists, nephrologists practicing outside the target countries, and nephrologists that do not routinely treat hemodialysis patients. Email addresses on each list were randomly allocated to either the Creatinine version or the eGFR version of the survey using a randomly permuted blocks stratified by country and generated independently of the email list manager.

Survey Administration

Surveys were distributed by email using Survey Monkey© (www.surveymonkey.com) and were completed online. Reminder emails were distributed twice after the initial distribution.

Before beginning the survey, respondents were asked to verify that they were practicing nephrologists and consented to the use of their results for research purposes. Eligible respondents were required to be practicing nephrologists in one of the target countries (Canada, the USA, the UK, Australia or New Zealand), and had to complete at least one question from one clinical vignette.

Outcomes

The primary outcome was the difference in the modified Likert scale scores between the two groups using the first three clinical vignettes (i.e. those that varied comorbidity and renal function). We assessed the subgroups of nephrologists by country of practice, number of years in practice ($<$ or \geq the median), type of clinical practice (focused on chronic kidney disease/ESRD, focused on transplantation, or both), and whether the respondent practiced at an academic or community hospital.

Statistical Analysis

Continuous variables were described as means and standard deviations (SD) or medians and interquartile ranges (25th to 75th percentile) and categorical variables were expressed as frequencies (%).

Differences in the modified Likert scale scores were calculated using random-effects regression to account for the multiple questions answered by each respondent. Respondents were considered a random intercept and the allocated survey version was considered a fixed effect. Initial modeling did not indicate improved model fit when respondents were treated as a random slope and the random slope component was dropped. *A priori* we evaluated whether there was an interaction between survey version and the level of eGFR for each question hypothesizing that

eGFR reporting would be less important at the extremes of renal function. There was a significant quantitative interaction at the $p < 0.001$ level for the lowest eGFR group (eGFR = 5 ml/min/1.73m²) compared to the highest eGFR group (eGFR = 12.5 ml/min/1.73m²) but not for any other groups. We therefore reported the primary outcome results separately for questions in which the eGFR was 5 ml/min/1.73m² while combining responses for all other eGFR values. We calculated the effect size as the mean difference between groups divided by the pooled standard deviation¹³. We conducted sensitivity analyses in which we converted the modified Likert scale scores to “Highly Likely” (score ≥ 6) or “Not Highly Likely” (score < 6) and used a multi-level logistic regression model to assess the relative odds of a “Highly Likely” score for the eGFR Group compared to the Creatinine Group.

The sample size was determined by convenience with the goal of obtaining a broad cross-section of nephrologists from each target country. We estimated that under the assumption of moderate autocorrelation, we would be able to detect a difference in scores of 0.7 with 90% power and a two-sided alpha of 0.025 if we received only 200 responses and we would be able to detect a difference of 0.3 if we received 1000 responses under the same assumptions (PASS version 11, Kaysville, Utah, USA).

All analyses were conducted with Stata version 11 MP (College Station, Texas, USA). A p-value < 0.05 was considered statistically significant.

