Developing a set of core outcomes for trials in hemodialysis: An international Delphi survey

Authors' full names and highest degree:

Nicole Evangelidis, BSocSc[1,2], Allison Tong, PhD[1,2], Braden Manns, MD, MSc[3], Brenda Hemmelgarn, MD, PhD[3], David C Wheeler, MD[4], Peter Tugwell, MD[5], Sally Crowe, PGDip[6], Tess Harris[7], Wim Van Biesen, MD, PhD[8], Wolfgang C Winkelmayer, MD, Sc.D[9], Benedicte Sautenet, MD, PhD[1,2,10,11], Donal O'Donoghue, MB[12], Helen Tam-Tham, MSc[3], Sajeda Youssouf, MD[12], Sreedhar Mandayam, MD, PhD[9], Angela Ju, BSc (Hons)[1,2], Carmel Hawley MBBS[13,14,15], Carol Pollock, PhD[16], David C Harris, MD, MBBS[17], David W Johnson, PhD[13,14,15], Dena E. Rifkin, MD[18], Francesca Tentori MD, MS[19], John Agar, MBBS[20], Kevan R Polkinghorne, PhD[21,22], Martin Gallagher, PhD[23], Peter G Kerr, PhD[21],Stephen P McDonald, PhD[24,25], Kirsten Howard, PhD[1], Martin Howell, PhD[1,2], Jonathan C Craig, PhD[1,2] for the Standardized Outcomes in Nephrology – Hemodialysis (SONG-HD) initiative[1]

Affiliations:

[1]Sydney School of Public Health, University of Sydney, Sydney, NSW, Australia

[2]Centre for Kidney Research, The Children's Hospital at Westmead, Sydney, NSW, Australia

[3]Departments of Medicine and Community Health Sciences; Libin Cardiovascular Institute and O'Brien Institute of Public Health, University of Calgary, Calgary, Canada

[4]Centre for Nephrology, University College London, London, United Kingdom

[5]Department of Medicine, University of Ottawa, Ottawa, Canada

[6]Crowe Associates, Oxon, United Kingdom

[7]PKD International, Geneva, Switzerland

[8]Renal Division, Ghent University Hospital, Ghent, Belgium

---

[1] http://songinitiative.org/

[9]Selzman Institute for Kidney Health, Section of Nephrology, Baylor College of Medicine, Houston, United States

[10]INSERM, U1153, Paris, France

[11]Department of Nephrology and Clinical Immunology, University Hospital of Tours, Tours, France

[12]Department of Renal Medicine, Salford Royal NHS Foundation Trust, Salford, United Kingdom

[13]Queensland School of Medicine, University of Queensland at Princess Alexandra Hospital, Brisbane, Australia

[14]Translational Research Institute, Brisbane, Australia

[15]Metro South and Ipswich Nephrology and Transplant Services (MINTS), Princess Alexandra Hospital, Brisbane, Australia

[16]Renal Division, Kolling Institute, Sydney, New South Wales, Australia

[17]Centre for Transplantation and Renal Research, The Westmead Institute for Medical Research, University of Sydney, Westmead, New South Wales, Australia.

[18]Division of Nephrology and Division of Preventive Medicine, University of California, San Diego

[19]Arbor Research Collaborative for Health, Ann Arbor, Michigan, United States

[20]University Hospital Geelong, Australia

[21]Monash Medical Centre and Monash University, Clayton, Australia

[22]Department of Epidemiology & Preventative Medicine, Monash University

[23]Concord Clinical School, The University of Sydney, Sydney, Australia

[24]Central Northern Adelaide Renal and Transplantation Service, Royal Adelaide Hospital, Adelaide, Australia

[25]Faculty of Health Science, University of Adelaide, Adelaide, Australia

Corresponding author:

Nicole Evangelidis

Centre for Kidney Research, The Children’s Hospital at Westmead, Westmead NSW 2145, Sydney, Australia

Tel: +61 2 9845 1474 | Fax: +61 2 9845 1491 | Email: nicole.evangelidis@sydney.edu.au

Word count (abstract):

297 words

Word count (body – excludes figures, tables and references):

3,492 words

Short title: Core outcomes in hemodialysis

ABSTRACT

Background: Survival and quality of life for patients on hemodialysis remain poor despite substantial research efforts. Existing trials often report surrogate outcomes that may not be relevant to patients and clinicians. The aim of this project was to generate a consensus-based, prioritized list of core outcomes for trials in hemodialysis.

Study Design: In a Delphi survey, participants rated the importance of outcomes using a 9-point Likert scale in round 1 and then re-rated outcomes in rounds 2 and 3 after reviewing other respondents' scores. For each outcome, the median, mean, and proportion rating 7-9 (critically important) were calculated.

Setting and Participants: 1,181 participants (202 [17%] patients/caregivers, 979 health professionals) from 73 countries completed round 1, with 838 (71%) completing round 3.

Outcomes and Measurements: Outcomes included in the potential core outcome set met the following criteria for both patients/caregivers and health professionals: median score ≥8; mean score ≥7.5; proportion rating the outcome 'critically important' ≥75% and median score <10 in the forced ranking question.

Results: Patients/caregivers rated four outcomes higher than health professionals: ability to travel (mean difference 0.9), dialysis-free time (0.5), dialysis adequacy (0.3), and washed out after dialysis (0.2). Health professionals gave a higher rating for mortality (1.0), hospitalization (1.0), drop in blood pressure (1.0), vascular access complications (0.9), depression (0.9), cardiovascular disease (0.8), target weight (0.7), infection (0.4) and potassium (0.4).

Limitations: The Delphi survey was conducted online in English-language and excludes participants without access to a computer and internet connection.

Conclusions: Patients/caregivers gave higher priority to lifestyle-related outcomes than health professionals. The prioritized outcomes for both groups were vascular access problems, dialysis adequacy, fatigue, cardiovascular disease and mortality. This process will inform a core outcome set that in turn will improve the relevance, efficiency and comparability of trial evidence to facilitate treatment decisions.

Keywords: hemodialysis, outcomes, Delphi survey, core outcome set, trials, outcome domains

INTRODUCTION

The enormous investment in biomedical research, particularly in randomized trials, may not have led to the improvements in health that were hoped for. [1-3] It has been estimated that 85% of the worldwide US $240 billion invested in research annually is wasted.[3] In nephrology, there has been substantial research investment into hemodialysis (HD), yet survival rates have not improved correspondingly over the past 40 years and quality of life remains poor even compared with patients with many cancers.[4-7] This may be partly attributable to what outcomes are selected and reported in trials – a challenge well-recognized across medical specialties.[1,8-10]

Surrogate endpoints are frequently used in clinical trials because of feasibility, in preference to outcomes that are directly relevant to patients and clinicians.[9,11,12] In HD, biochemical markers such as serum phosphorus, calcium, and parathyroid hormone, are commonly reported but are not strongly and consistently associated with mortality, cardiovascular disease (CVD), or quality of life.[13-17] Patients on HD prioritize outcomes relevant to their well-being and lifestyle – fatigue, ability to travel, ability to work, sleep, anxiety/stress [18,19] – all largely absent as outcomes reported by HD trials. In addition, the large heterogeneity of outcome measures and potential for outcome reporting bias (where trials selectively report results for outcomes that favor the intervention) undermines the reliability of trial evidence to inform clinicians and patients about the relative effects of interventions.[20]

Engaging all stakeholders in establishing a core outcome set, an agreed minimum set of standardized outcomes to be measured and reported in all trials for a specific clinical area [21,22], can increase the relevance, efficiency and reliability of trials. Initiatives to develop core outcomes are seen in rheumatology and oncology, and have demonstrated improvements in consistent reporting of relevant outcomes.[10,23,24] As part of the international Standardized Outcomes in Nephrology-Hemodialysis (SONG-HD) initiative, this study aimed to generate a consensus-based prioritized list of outcome domains for people on HD, which will be used to establish a core outcome set that reflects the shared priorities of patients, caregivers and health professionals.

METHODS

Study design

The Delphi method is a technique for achieving consensus among a panel of experts. This process involves sequential surveys, typically conducted over three rounds, answered anonymously and gives equal influence to all who participate. It was first developed by the RAND Corporation in the 1950's [25] and has since been increasingly used as a valid approach to develop consensus-based core outcomes for clinical trials in various medical specialty areas.[10,22,26-28] The SONG-HD Delphi process is shown in Supplementary Figure S1.

Participant selection and recruitment

Stakeholders including patients, caregivers/family members, nephrologists, surgeons, nurses, social workers, psychologists, dieticians, pharmacists, policy makers, researchers and industry, with experience or interest in HD were invited to join the Delphi Panel. Participants worldwide were eligible if they were aged over 18 years and able to complete an online survey in English-language. All participants provided informed consent.

Using an opt-in, snowballing sampling frame, we recruited patients/caregivers through participating hospitals, patient/consumer organizations, and social media listed in Supplementary File S1. Health professionals were recruited via the investigators networks and via emails and newsletters circulated by professional societies (Supplementary File S1). Participants registered their email on www.songinitiative.org prior to the survey launch. The ethics boards of the University of Sydney (2015/228), Baylor College of Medicine (H-37406), University of Calgary (REB15-0708), Monash Medical Centre (13082B), Salford Royal NHS (15/WM/0303), and Sydney West Area Health Service (HREC2009/6/4.15) approved this study.

Data collection

The 34 outcome domains for the three-round Delphi survey were identified from a systematic review of outcomes reported in trials in HD, stakeholder interviews, and nominal group technique conducted with patients on HD and caregivers.[29,30] The ordering of outcomes was randomized and included a plain language definition (Supplementary File S2). The survey was reviewed by the SONG Executive Committee and SONG-HD investigators and piloted among 10 patients. The Delphi survey was completed online via LimeSurvey between September and November, 2015. The online survey administration minimizes data entry error, allows for wider dissemination and is more efficient compared to a paper survey.

Round 1: Participants rated the importance of each of the 34 outcomes based on a 9-point Likert scale. A score of 7-9 indicated that the outcome was of "critical importance", 4-6 indicated "important but not critical" and 1-3 indicated "limited importance" according to the GRADE process.[31] An option of "unsure" was provided. Participants could enter comments about their choice of ranking for each outcome. In addition, participants could suggest new outcomes that were not included in the survey. Outcomes with a mean and median of less than 7 for patients/ caregivers and health professionals were not included in round 2.

Round 2: Participants reviewed the group scores and their own score for each outcome and re-rated the 29 outcomes using the same 9-point Likert scale. The group scores were displayed in an interactive column graph which showed the distribution of scores for: patients/caregivers, health professionals, and the total sample combined (weighted) (Supplementary Figure S1). Instructions on how to read the graph were provided to ensure that participants were able to understand the results. For each outcome, an optional comments box allowed participants to explain reasons for their rating. Outcomes with a mean and median ≤ 7 for patients/ caregivers and health professionals were excluded from round 3.

Round 3: Participants were asked to re-rate 20 outcomes using the same Likert scale in the previous rounds after viewing the scores, and in addition, de-identified comments (i.e. free text responses from participants relating to reasons for their rankings or observations on the results for each outcome) from round 2. The comments were divided into two boxes: ‘Patients and caregivers’ and ‘Health professionals’ with the ability to scroll down and read all comments. A free text box was provided for each outcome so participants could provide additional comments. In addition, participants completed a forced ranking question, using a drag and drop function, to rank outcomes relative to each other.

Data analysis

We used SPSS (IBM; Version 22.0) to calculate descriptive statistics. We calculated the median, mean, and proportion of participants (rating 7-9) for each outcome. The scores were calculated separately for patients/caregivers and health professionals, with the difference in means considered significant at $P<0.05$ based on the t-test. For the ranking scores, we calculated the median and interquartile range (IQR) for each outcome to determine rank. Any analysis of the total sample was weighted equally between patients/caregivers and health professionals.

Consensus was defined a priori based on the OMERACT definition using proportion scores. "Consensus in" is defined as greater than or equal to 70% of participants scoring as 7 to 9 and less than 15% participants scoring as 1 to 3.[29] However, as most participants rated all outcomes in round 3 as critically important with scores of 7-9, these criteria resulted in a list of 16 outcomes, which exceeded the recommended 3 to 5 outcomes for a core outcome set. Therefore, the definition and threshold for "consensus in" were revised to determine a maximum of five core outcomes to be considered for the core outcome set.

Definition of consensus

Outcomes from round 1 with a mean and median score greater than or equal to 7 for patients/caregivers and health professionals were included in round 2. This was validated against the proportion of critically important scores (rated 7-9) for each outcome to ensure that important outcomes were not excluded.

Outcomes from round 2 with a mean and median score greater than 7 for patients/caregivers and health professionals were included in round 3. This was validated against the proportion of critically important scores (rated 7-9) for each outcome.

Outcomes included in the potential core outcome set met the following criteria for both patients/caregivers and health professionals: median score greater than or equal to 8; mean score greater than or equal to 7.5; proportion of participants rating the outcome 'critically important' is greater than or equal to 75% and median score is less than 10 in the forced ranking question.

RESULTS

Participant characteristics

In total, 1 181 people from 73 countries participated in round 1 of the Delphi survey including 202 (17%) patients/caregivers and 979 (83%) health professionals. Round 2 included 165 patients/caregivers (17%) and 784 health professionals (83%) from 63 countries. In the third and final round, 150 patients/caregivers (18%) and 688 health professionals (82%) participated. The full survey completion rate was 71%. The participant characteristics are provided in Tables 1 and 2.

In round 3, 115 (77%) patients/caregivers were aged from 41 to 70 years and 76 (51%) were women. Patients/caregivers were from 14 countries in round 1 (11 countries in round 3). In round 3, the majority of patient/caregiver participants were from Australia (40 [27%]), Canada (37 [24.7%]), United Kingdom (UK) (25 [17%]), United States (US) (19 [13%]), and New Zealand (18 [12%]). Among the 116 (77%) patients on HD, 63 (42%) patients were on in-center HD and 51 (34%) were on home HD. The health professionals included 857 nephrologists (51%), 386 nurses (38%), 53 researchers (5%), and 63 (6%) in other roles. Health professionals were from 72 countries in round 1 (62 countries in round 3).

Delphi scores

Round 1: The means, medians, and proportion of participants rating the outcome 7 – 9 (critical importance) for each of the 34 outcomes are shown in Supplementary Table S1. The top three outcomes rated by patients/caregivers based on mean scores (1 to 9) were: dialysis adequacy (7.5 [standard deviation [SD] =2.1]), ability to travel (7.5 [SD 1.9]) and dialysis-free time (7.3 [SD 1.8]). The top three outcomes for health professionals were: vascular access problems (8.1 [SD 1.3]), CVD (7.9 [SD 1.3]), and mortality (7.7 [SD 1.6]. Definitions of high rating outcomes are provided in Box 1.

The following outcomes were excluded from round 2 as they had a mean or median score of less than 7 (not of critical importance) among both patient/caregiver and health professional groups in round 1: nausea/vomiting, sexual function, restless legs syndrome, itching and cramps. Less than 25% of participants suggested new outcomes in round 1 (Supplementary Table S2). These outcomes were not considered for inclusion in round 2 due to the following reasons: the outcome could not be measured in a clinical trial for the majority of adult patients on HD, the outcome was too broad conceptually or ambiguously defined, and the outcome was described as an intervention.

Round 2: Round 2 included 29 outcomes (Supplementary Table S3). The top 3 outcomes for patients/caregivers were: dialysis adequacy (7.7 [SD 1.8]), ability to travel (7.6 [SD 1.9]) and dialysis-free time (7.5 [SD 1.7]). The top 3 outcomes rated by health professionals were: vascular access problems (8.4 [SD 1.0]), CVD (8.2 [SD 1.1]) and mortality (8.2 [SD 1.2]).

Outcomes which had a mean and median score ≤7 with less than 70% of the sample rating the outcome 7-9 (critical importance) were excluded from round 3: anxiety/stress, food enjoyment, calcium, parathyroid hormone, cognition, sleep, bone health, financial impact and phosphate.

Round 3: Round 3 included 20 outcomes (Supplementary Table S4). The top three outcomes for patients/caregivers were: dialysis adequacy (7.9 [SD 1.8]), ability to travel (7.7 [SD 1.7]) and vascular access problems (7.7 [SD 2.0]). The top three outcomes for health professionals were vascular access problems (8.6 [SD 0.9]), CVD (8.4 [SD 1.0]) and mortality (8.3 [SD 1.1]).

Outcomes that met at least two of the following criteria for consensus within both stakeholder groups (median ≥ 8, mean ≥ 7.5, proportion ≥ 75%, and median rank < 10) [Table 3]) were: CVD, mortality, dialysis adequacy, fatigue, and vascular access problems. All participant comments for each outcome are provided in Supplementary File S3.

Changes in scores from round 1 to 3 within stakeholder groups

As shown in Figure 1, the patient/caregiver mean scores increased between rounds 1 and 3 for the following 7 outcomes: vascular access problems (mean score difference 0.9, P = <0.001), CVD (0.7, P = 0.002), infection/immunity (0.7, P = 0.004), drop in blood pressure (0.6, P = 0.02), mobility (0.6, P = 0.02), target weight (0.6, P = 0.02) and washed out after dialysis (0.5, P = 0.01).

For health professionals, the mean scores increased for 16 outcomes between rounds 1 and 3 (Figure 2): mortality (0.6, P <0.001), CVD (0.5, P < 0.001), vascular access problems (0.4, P < 0.001), ability to travel (0.4, P < 0.001), dialysis-free time (0.4, P < 0.001), dialysis adequacy (0.4, P < 0.001), washed out after dialysis (0.3, P <0.001), ability to work (0.3, P < 0.001), infection/immunity (0.3, P < 0.001), drop in blood pressure (0.3, P < 0.001), hospitalization (0.3, P < 0.001), fatigue (0.3, P < 0.001), impact on family/friends (0.3, P < 0.001), mobility (0.2, P < 0.001), pain (0.2, P = 0.007) and blood pressure (0.2, P = 0.02). No mean scores decreased across the three rounds of the survey for either patients/caregivers or health professionals.

Differences between stakeholder groups

The differences in ratings between stakeholder groups are shown in Figure 3. Based on the difference in mean scores in round 3, four outcomes were rated higher by patients/caregivers: ability to travel (mean difference 0.9, P <0.001), dialysis-free time (0.5, P <0.001), dialysis adequacy (0.3, P = 0.05) and washed out after dialysis (0.2, P = 0.05).

Health professionals rated 11 outcomes higher compared to patients/caregivers: mortality (mean difference 1.0, P <0.001), hospitalization (1.0, P <0.001), drop in blood pressure (1.0, P <0.001), vascular access problems (0.9, P <0.001), depression (0.9, P <0.001), CVD (0.8, P <0.001), target weight (0.7, P <0.001), infection/immunity (0.4, P = 0.002), potassium (0.4, P = 0.02), ability to work (0.3, P = 0.008), and pain (0.3, P = 0.04).

Forced ranking scores

The results of the forced ranking question are shown in Supplementary Table S5. The top outcomes ranked by patients/caregivers were (median rank score [Inter Quartile Range (IQR)]): CVD (7 [IQR 4-14]), vascular access problems (median rank 7 [IQR 3-13]), dialysis adequacy (8[IQR 4-11]) and fatigue (8 [IQR 4-11]). The top outcomes ranked by health professionals were: CVD (3 [IQR 2-7]), mortality (3 [IQR 1-9]) and vascular access problems (4 [IQR 2-7]).

DISCUSSION

The highest-priority outcomes shared among patients/caregivers and health professionals were vascular access problems, dialysis adequacy, fatigue, CVD, and mortality. Overall, most of these reflect common and high-impact outcomes in the context of HD, which have remained as major challenges in providing care for patients on HD. Frequently reported biochemical outcomes in HD trials, such as phosphate, calcium and parathyroid hormone, were consistently rated to be of lower importance by both stakeholder groups.

Although there was convergence in ratings between patients/caregivers and health professionals across the three rounds of the Delphi, our findings also highlight some mismatches. Outcomes relating to lifestyle (ability to travel, dialysis-free time) and well-being (washed out after dialysis) rated higher among patients/ caregivers compared to health professionals. Research has consistently shown that patients and caregivers prioritize lifestyle-related outcomes over biochemical endpoints.[18,19,30] Ability to travel, fatigue/energy and dialysis-free time are important outcomes for patients and caregivers who want to retain normality in their day to day lives and maintain a reasonable quality of life.[19,30] In a recent discrete choice study, patients were willing to trade off 23 months of life expectancy with home-based dialysis to increase their ability to travel.[32] Biochemical endpoints are less important to patients and caregivers because they are considered intangible or imperceptible.[30]

Dialysis adequacy was also rated higher by patients/caregivers. The participant comments (Supplementary File S3) and discussions from the recent SONG-HD consensus workshop on establishing core outcomes in HD[33], indicates participants conceptualized the term “dialysis adequacy” as a broad quality of life outcome (i.e. dialysis that is adequate for enabling patients to feel well) rather than quantification of urea kinetics, which may explain this apparent divergence.

The rating of CVD by patients increased significantly between round 1 and 3. Based on the comments provided by patients shown in Supplementary File S3, the reasons were largely due to gaining an understanding of the importance of CVD through reading other participant’s comments, and a few patients who experienced cardiovascular events during the survey period (Supplementary File S3). Similarly, studies have shown that patients with chronic kidney disease have limited knowledge about their risk of CVD.[30,34]

The outcomes rated higher by health professionals (with a mean difference greater than 0.5 in ratings) were: mortality, hospitalization, drop in blood pressure, vascular access problems, depression, CVD, and target weight. This perhaps reflects their perceived primary clinical role in preventing and managing these outcomes which are common among HD patients and awareness of their impact on patients' lives, whereas high-priority outcomes for patients such as ability to travel and dialysis-free time may be seen by health professionals as impractical to measure.

Uremic symptoms such as itching and nausea/vomiting, as well as restless legs syndrome and cramps were rated relatively lower by patients/caregivers in round 1 and were excluded from subsequent rounds. Instead, patients/caregivers gave higher priority to broader outcomes related to lifestyle impact and overall well-being (dialysis-free time, ability to travel, fatigue) rather than specific symptoms. Yet, many patient-reported outcome measures are designed to assess symptoms and few robust and well-validated instruments measure specific patient-important outcomes that have an arguably more pervasive and long-term impact on lifestyle and well-being.[35-37] Furthermore, patients may not report lifestyle problems (e.g. ability to work) which are not articulated in clinical settings as patients may perceive these to be beyond their clinician's control.

The SONG-HD Delphi survey had a large sample size (five times greater than most similar Delphi surveys to establish core outcomes)[23,38,39] and achieved a broad engagement of a range of stakeholders including patients/caregivers, multidisciplinary healthcare providers, policy makers and industry, across 73 countries, and retained a high response rate of 71% by round 3. The process provided a transparent and systematic way to gain consensus on the importance of outcomes to consider for a core outcome set in HD. However, there are some potential limitations. The Delphi survey was conducted online and excludes participants without access to a computer and internet connection. The survey was only available in English-language to retain consistency of meaning and for feasibility, although some participants submitted open-text responses in Spanish-language which were translated. We also acknowledge that detailed analysis of the open-text responses was not conducted and is beyond the scope of the current paper. Given the design of the Delphi survey, we recognize that the results are potentially biased towards participants who are English speaking with access to a computer and internet connection, and who have the ability to use a computer.

We included participants from 73 countries and acknowledge that the majority of participants were from high income countries. We recognize that these outcomes may not be shared by those who did not participate in the study, or by individuals or groups within the study. A study of this type must necessarily accept the views of the majority of participants as representing the consensus position, and will not represent the view of all individuals.

Whilst definitions were provided for each outcome domain, we acknowledge the inevitable interaction between some outcome domains, and participants may have interpreted the outcomes differently. We provided the participant comments to make explicit how the outcomes were conceptualized, and based on these data, we did not note any apparent differences that may have explained variation in prioritization between patients/caregivers and health professionals. The results from the consensus workshop further support this.[33]

Currently, there is no core outcome set in chronic kidney disease. The Delphi survey results will be used to establish a core outcome set to be reported in all trials in HD, which is expected to increase the quality and relevance of research. This has been done successfully by the Outcome Measures in Rheumatology (OMERACT) initiative, which has improved the reporting and relevance of outcomes in rheumatology trials.[10,24] Recent analyses have shown that a higher percentage of trials are now incorporating the OMERACT core outcome set in rheumatology trials.[24]

The results of this study were discussed at a recent SONG-HD consensus workshop. To effectively implement a core outcome set in chronic kidney disease, patients/caregivers and health professional participants suggested that outcome measures should be feasible and valid, and take patient priorities into account.[33] Once the core outcome domains have been established, which will be largely informed by the results of this Delphi survey, further work will focus on identifying core outcome measures for the outcome domains.

In summary, the international SONG-HD Delphi study identified a prioritized set of outcome domains for trials in HD based on consensus among key stakeholder groups. The top prioritized outcomes by both patients/caregivers and health professionals were vascular access problems, dialysis adequacy, fatigue, CVD and mortality. Patients/caregivers place highest value on outcomes that will enable them to maintain their day-to-day well-being and lifestyle. The findings will directly inform the development of a core outcome set to be used in HD trials. Ultimately, this will strengthen the relevance and reliability of trial evidence to support shared decision-making for people dependent on HD in order to improve treatment outcomes.

ACKNOWLEDGEMENTS

We thank the following organizations for their support:
International: Cochrane Kidney and Transplant, International Society of Nephrology (ISN), Kidney Disease | Improving Global Outcomes (KDIGO), PKD International; Australia/New Zealand: Australian and New Zealand Society of Nephrology (ANZSN), Australian Kidney Trials Network (AKTN), Christchurch Kidney Society, Kidney Health Australia (KHA), Kidney Health Australia – Caring for Australasians with Renal Impairment Guidelines (KHA-CARI), Kidney Health New Zealand (KHNZ), Renal Society of Australasia (RSA); Canada: The Kidney Foundation of Canada (KFOC), Canadian Society of Nephrology (CSN); Europe: British Kidney Patient Association, British Renal Society (BRS), European Kidney Patients Federation (EKPF), European Renal Best Practice (ERBP), The Renal Association; United States: Home Dialysis Central, National Kidney Foundation Southern California

Contributions

Research idea and study design: all authors; Data acquisition: NE, AT, BM, BH, DCW, PT, SC, TH, WvB, WCW, HT, SY, SM, DCH, JCC; Data analysis/interpretation: all authors. Statistical analysis: NE, BS, JCC. Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved. NE takes responsibility that this study has been reported honestly, accurately, and transparently; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Support

SONG-HD is funded by the NHMRC (1098815). Dr Tong is supported by an NHMRC Fellowship

(1106716). The funding organization had no role in the design and conduct of the study; collection; management, analysis and interpretation of the data; or preparation, review, or approval of the manuscript.

Financial Disclosure

David C Harris received fees from Roche Myanmar to give a lecture in Myanmar. David W Johnson has received consultancy fees, research grants, travel sponsorships and speaker's honoraria from Baxter Healthcare and Fresenius Medical Care. He has also received travel sponsorships from Amgen. Sreedhar Mandayam has received research funding from Keryx and is on the Advisory Board/consulting for Corp Medix. Carol Pollock has received lecture/consulting fees from Astra Zeneca, MSD, Boehringer Ingelheim and Baxter Healthcare. David C Wheeler has received consulting fees from Amgen, Boehringer Ingelheim, Akebia, UCB Celltech, Bristol Myers Squibb, Vifor Pharma, Otsuka and Janssen. He has received lecture fees from Fresenius, Amgen, Janssen, ZS Pharma and Vifor Fresenius. All other authors have no financial interests to disclose.

REFERENCES

1. Chalmers I, Bracken M, Djulbegovic B, Garattini S, Grant J, Gulmezoglu A. How to increase value and reduce waste when research priorities are set. *Lancet.* 2014;383:156 - 165.

2. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet Lond Engl.* Jul 4 2009;374(9683):86-89.

3. Macleod MR, Michie S, Roberts I, et al. Biomedical research: increasing value, reducing waste. *Lancet.* Jan 11 2014;383(9912):101-104.

4. Bell CM, Chapman RH, Stone PW, Sandberg EA, Neumann PJ. An off-the-shelf help list: a comprehensive catalog of preference scores from published cost-utility analyses. *Med Decis Making.* Jul-Aug 2001;21(4):288-294.

5. Nissenson A. Improving outcomes for ESRD patients: shifting the quality paradigm. *Clin J Am Soc Nephrol.* 2014;9:430 - 434.

6. Wong G, Howard K, Chapman J, Pollock C, Chadban S, Salkeld G. How do people with chronic kidney disease value cancer-related quality of life? *Nephrology.* 2012;17:32 - 41.

7. Wyld M, Morton R, Hayen A, Howard K, Webster A. A systematic review and meta-analysis of utility-based quality of life in chronic kidney disease treatments. *PLoS Med.* 2012;9:e1001307.

8. Gargon E, Gurung B, Medley N, Altman D, Blazeby J, Clarke M. Choosing important health outcomes for comparative effectiveness research: a systematic review. *PLoS One.* 2014;9:e99111.

9. Ghandi G. Patient-important outcomes in registered diabetes trials. *JAMA.* 2008;299:2543 - 2549.

10. Boers M, Kirwan J, Wells G, Beaton D, Gossec L, d'Agostino M. Developing core outcome measurement sets for clinical trials: OMERACT filter 2.0. *J Clin Epidemiol.* 2014;67:745 - 753.

11. Svensson S, Menkes D, Lexchin J. Surrogate outcomes in clinical trials: a cautionary tale. *JAMA Intern Med.* 2013;173:612 - 612.

12. Yudkin J, Lipska K, Montori V. The idolatory of the surrogate. *BMJ.* 2011;343:d7995.

13. Halimi JM, Sautenet B, Gatault P, Roland M, Giraudeau B. Renal endpoints in renal and cardiovascular randomized clinical trials: time for a consensus? *Fundamental & clinical pharmacology.* Dec 2012;26(6):771-782.

14. Investigators TET. Effect of Cinacalcet on Cardiovascular Disease in Patients Undergoing Dialysis. *N Engl J Med.* 2012;367(26):2482-2494.

15. Palmer SC, Hayen A, Macaskill P, et al. Serum levels of phosphorus, parathyroid hormone, and calcium and risks of death and cardiovascular disease in individuals with chronic kidney disease: a systematic review and meta-analysis. *JAMA.* Mar 16 2011;305(11):1119-1127.

16. Palmer SC, Teixeira-Pinto A, Saglimbene V, et al. Association of Drug Effects on Serum Parathyroid Hormone, Phosphorus, and Calcium Levels With Mortality in CKD: A Meta-analysis. *Am J Kidney Dis.* Dec 2015;66(6):962-971.

17. Perkovic V, Neal B. Trials in kidney disease--time to EVOLVE. *N Engl J Med.* Dec 27 2012;367(26):2541-2542.

18. Morton R, Tong A, Howard K, Snelling P, Webster A. The views of patients and carers in treatment decision making for chronic kidney disease: systematic review and thematic synthesis of qualitative studies. *BMJ.* 2010;340:c112.

19. Morton R, Tong A, Webster A, Snelling P, Howard K. Characteristics of dialysis important to patients and family caregivers: a mixed methods approach. *Nephrol Dial Transplant.* 2011;26:4038 - 4046.

20. Saini P, Loke Y, Gamble C, Altman D, Williamson P, Kirkham J. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ.* 2014;349:g6501.

21. Clark M. Standardising outcomes for clinical trials and systematic reviews. *Trials.* 2007;8:39.

22. Prinsen C, Vohra S, Rose M, King-Jones S, Ishaque S, Bhaloo Z. Core Outcome Measures in Effectiveness Trials (COMET) initiative: protocol for an international Delphi study to achieve consensus on how to select outcome measurement instruments for outcomes included in a 'core outcome set'. *Trials.* 2014;15:247.

23. Gorst SL, Gargon E, Clarke M, Blazeby JM, Altman DG, Williamson PR. Choosing Important Health Outcomes for Comparative Effectiveness Research: An Updated Review and User Survey. *PloS one.* 2016;11(1):e0146444.

24. Kirkham J, Boers M, Tugwell P, Clarke M, Williamson P. Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. *Trials.* 2013;14:324.

25. Dalkey N, Helmer O. An Experimental Application of the DELPHI Method to the Use of Experts. *Management Science.* 1963;9(3):458-467.

26. Chiarotto A, Terwee C, Deyo R, Boers M, Lin C, Buchbinder R. A core outcome set for clinical trials on non-specific low back pain: study protocol for the development of a core domain set. *Trials.* 2014;15:511.

27. Harman N, Bruce I, Callery P, Tierney S, Sharif M, O'Brien K. MOMENT--Management of Otitis Media with Effusion in Cleft Palate: protocol for a systematic review of the literature and identification of a core outcome set using a Delphi survey. *Trials.* 2013;14:70.

28. Waters A, Tudur Smith C, Young B, Jones T. The CONSENSUS study: protocol for a mixed methods study to establish which outcomes should be included in a core outcome set for oropharyngeal cancer. *Trials.* 2014;15:168.

29. Tong A, Manns B, Hemmelgarn B, et al. Standardised outcomes in nephrology - Haemodialysis (SONG-HD): study protocol for establishing a core outcome set in haemodialysis. *Trials.* 2015;16(1):364.

30. Urquhart-Secord R, Craig JC, Hemmelgarn B, et al. Patient and Caregiver Priorities for Outcomes in Hemodialysis: An International Nominal Group Technique Study. *Am J Kidney Dis.* Sep 2016;68(3):444-454.

31. Schunemann H, Brozek J, Oxman A. *GRADE Handbook for grading quality of evidence and strength of recommendation.* 2009.

32. Morton RL, Snelling P, Webster AC, et al. Dialysis Modality Preference of Patients With CKD and Family Caregivers: A Discrete-Choice Study. *Am J Kidney Dis.* 7// 2012;60(1):102-111.

33. Tong A, Manns B, Hemmelgarn B, et al. Establishing Core Outcome Domains in Hemodialysis: Report of the Standardized Outcomes in Nephrology-Hemodialysis (SONG-HD) Consensus Workshop. *Am J Kidney Dis.* Aug 3 2016.

34. Lopez-Vargas PA, Tong A, Howell M, et al. Patient awareness and beliefs about the risk factors and comorbidities associated with chronic kidney disease - a mixed-methods study. *Nephrology.* Jun 1 2016.

35. Breckenridge K, Bekker HL, Gibbons E, et al. How to routinely collect data on patient-reported outcome and experience measures in renal registries in Europe: an expert consensus meeting. *Nephrol Dial Transplant.* October 1, 2015 2015;30(10):1605-1614.

36. Flythe JE, Powell JD, Poulton CJ, et al. Patient-Reported Outcome Instruments for Physical Symptoms Among Patients Receiving Maintenance Dialysis: A Systematic Review. *Am J Kidney Dis.* Dec 2015;66(6):1033-1046.

37. Purnell TS, Auguste P, Crews DC, et al. Comparison of life participation activities among adults treated by hemodialysis, peritoneal dialysis, and kidney transplantation: a systematic review. *Am J Kidney Dis.* Nov 2013;62(5):953-973.

38. Chiarotto A, Deyo RA, Terwee CB, et al. Core outcome domains for clinical trials in non-specific low back pain. *Eur Spine J.* Jun 2015;24(6):1127-1142.

39. Harman N, Bruce I, Kirkham J, et al. The Importance of Integration of Stakeholder Views in Core Outcome Set Development: Otitis Media with Effusion in Children with Cleft Palate. *PloS one.* 2015;10(6):e0129514.

TABLES AND FIGURES

Table 1. Characteristics of patients/caregivers

Table 2. Characteristics of health professionals

Box 1. Definitions of high rating outcome domains

Table 3. Inclusion criteria for the core outcome set based on median, mean, proportion (7-9) and median rank scores

Figure 1. Mean scores of patients/caregivers in rounds 1-3

Note: Ordered by round 3 scores. Round 1 (n=202); round 2 (n=165); round 3 (n=150) Mean scores for round 2 and 3 are not available for outcomes that were excluded in those rounds.

Figure 2. Mean scores of health professionals in rounds 1-3

Note: Ordered by round 3 scores. Round 1 (n=979); round 2 (n=784); round 3 (n=688). Mean scores for round 2 and 3 are not available for outcomes that were excluded in those rounds.

Figure 3. Difference in mean scores between patients/caregivers and health professionals (rounds 1-3)

Note: Ordered by difference in mean scores between patients/caregivers and health professionals. Error bars refer to standard deviations.

SUPPLEMENTARY MATERIAL

Figure S1. SONG-HD Delphi process

File S1. Collaborating organizations

File S2. Survey outcome definitions

Table S1. Round 1 outcome scores (mean, median, proportion 7-9 %) of patients/caregivers and health professionals

Table S2. Suggested outcomes by participants (n=1 181) in round 1

Table S3. Round 2 outcome scores (mean, median, proportion 7-9 %) of patients/caregivers and health professionals

Table S4. Round 3 outcome scores (mean, median, proportion 7-9 %) of patients/caregivers and health professionals

Table S5. Median rank scores, forced ranking question (round 3)

File S3. Participant comments Delphi survey (rounds 1-3)

Figure S2. Median scores of patients/caregivers in rounds 1 -3

Note: Ordered by round 3 scores. Round 1 (n=202); round 2 (n=165); round 3 (n=150). Median scores for round 2 and 3 are not available for outcomes that were excluded in those rounds.

Figure S3. Median scores of health professionals in rounds 1-3

Note: Ordered by round 3 scores. Round 1 (n=979); round 2 (n=784); round 3 (n=688). Median scores for round 2 and 3 are not available for outcomes that were excluded in those rounds.

Table 1. Characteristics of patients/caregivers

Characteristic		n (%)	
	Round 1 **n=202**	**Round 2** **n=165**	**Round 3** **n=150**
Participant type			
Patient	168 (83.2)	138 (83.6)	127 (84.7)
Caregiver/family member	34 (16.8)	27 (16.4)	23 (15.3)
Gender			
Male	96 (47.5)	78 (47.3)	74 (49.3)
Female	106 (52.5)	87 (52.7)	76 (50.7)
Age group (years)			
18 – 40	33 (16.4)	23 (13.9)	16 (10.7)
41 – 50	41 (20.3)	34 (20.6)	29 (19.3)
51 – 60	59 (29.2)	49 (29.7)	47 (31.3)
61 – 70	48 (23.8)	40 (24.2)	39 (26.0)
≥ 71	21 (10.4)	19 (11.5)	19 (12.7)
Marital status*			
Single	24 (11.9)	20 (12.1)	17 (11.3)
Partner/de-facto	14 (7.0)	11 (6.7)	9 (6.5)
Married	110 (54.5)	87 (52.7)	78 (52.0)
Divorced/separated/widowed	41 (20.3)	34 (20.6)	34 (24.6)
Number of children*			
0	64 (31.7)	51 (30.9)	43 (28.7)
1-2	83 (41.1)	66 (40.0)	63 (42.0)
3-4	41 (20.3)	34 (20.6)	31 (20.7)
Employment status*			
Employed	74 (36.7)	55 (33.3)	46 (35.4)
Unemployed	37 (18.3)	28 (17.0)	25 (16.7)
Retired	67 (33.2)	60 (36.4)	58 (38.7)
Student	1 (0.5)	1 (0.6)	1 (0.8)
Education*			
Did not complete high school	28 (13.9)	21 (12.7)	18 (12.0)
High school graduate	29 (14.4)	22 (13.3)	22 (14.7)
Professional certificate	43 (21.3)	36 (21.8)	32 (21.3)
Undergraduate degree	59 (29.2)	47 (28.5)	43 (28.7)
Postgraduate degree	24 (11.9)	22 (13.3)	19 (12.7)
Current type of treatment*			
In center hemodialysis	91 (45.0)	71 (43.0)	63 (42.0)
Home hemodialysis	70 (34.7)	57 (34.5)	51 (34.0)
Peritoneal dialysis	2 (1.0)	2 (1.2)	2 (1.3)
Transplant	19 (9.4)	17 (10.3)	17 (11.3)
Years on hemodialysis*			
< 1	26 (12.9)	18 (10.9)	17 (11.3)
1-5	95 (47.0)	76 (46.1)	69 (46.0)
6-10	27 (13.4)	25 (15.2)	22 (14.7)
11-15	20 (9.9)	17 (10.3)	14 (9.3)
> 15	15 (7.4)	12 (7.3)	12 (8.0)

Country			
Canada	53 (26.2)	41 (24.8)	37 (24.7)
Australia	49 (24.3)	41 (24.8)	40 (26.7)
United Kingdom	35 (17.3)	29 (17.6)	25 (16.7)
United States	25 (12.4)	20 (12.1)	19 (12.7)
New Zealand	21 (10.4)	20 (12.1)	18 (12.0)
Other[a]	19 (9.5)	14 (8.4)	11(7.3)

*Percentages do not add to 100 due to undisclosed responses (excluded).

[a] Other includes 9 countries: Romania, India, Spain, Czech Republic, Egypt, Netherlands, Indonesia, Italy and Philippines.

Table 2. Characteristics of health professionals

Characteristic	**n (%)** **Round 1** **n=979**	**Round 2** **n=784**	**Round 3** **n=688**
Participant type[a]			
Nephrologist	483 (46.9)	450 (57.4)	401 (58.3)
Nurse	386 (37.5)	277 (35.3)	233 (33.9)
Researcher	53 (5.2)	50 (6.4)	47 (6.8)
Nephrology trainee	44 (4.3)	36 (4.3)	28 (3.8)
Policy maker	17 (1.7)	16 (1.9)	16 (2.2)
Industry	13 (1.3)	13 (1.6)	12 (1.6)
Dietician	11 (1.1)	9 (1.1)	9 (1.2)
Social worker	7 (0.7)	7 (0.8)	7 (1.0)
Pharmacist	4 (0.4)	3 (0.4)	3 (0.4)
Psychologist	3 (0.3)	1 (0.1)	1 (0.1)
Surgeon	2 (0.2)	2 (0.2)	2 (0.3)
Other	6 (0.6)	3 (0.4)	3 (0.4)
Gender			
Male	447 (45.7)	362 (46.2)	318 (46.2)
Female	532 (54.3)	422 (53.8)	370 (53.8)
Age group (years)			
18 - 40	435 (44.4)	317 (40.4)	268 (39.0)
41 – 50	262 (26.8)	220 (28.1)	197 (28.6)
51 – 60	207 (21.1)	181 (23.1)	164 (23.8)
61 – 70	63 (6.4)	58 (7.4)	52 (7.6)
≥ 71	12 (1.2)	8 (1.0)	7 (1.0)
Experience in HD (years)			
≤10	406 (41.5)	297 (37.9)	252 (36.6)
11-20	308 (31.5)	254 (32.4)	228 (33.1)
≥ 21	265 (27.1)	117 (29.7)	63 (9.1)
No. of hemodialysis trials as investigator			
0	439 (44.8)	336 (42.9)	287 (41.7)
1-5	364 (37.2)	299 (38.1)	266 (38.7)
6-10	93 (9.5)	80 (10.2)	72 (10.5)
≥ 11	83 (8.5)	69 (8.8)	63 (9.2)
Other roles			
Government, policy	119 (10.9)	93 (10.6)	87 (11.2)
Guidelines	391 (35.8)	314 (35.6)	278 (35.6)
Funding	73 (6.7)	66 (7.5)	60 (7.7)
Other	509 (46.6)	408 (46.3)	355 (45.5)
Country			
Australia	133 (13.6)	120 (15.3)	108 (15.7)
Saudi Arabia	131 (13.4)	83 (10.6)	66 (9.6)
Spain	122 (12.5)	98 (12.5)	85 (12.4)
Romania	101 (10.3)	73 (9.3)	59 (8.6)
Canada	58 (5.9)	48 (6.1)	42 (6.1)
Portugal	56 (5.7)	48 (6.1)	44 (6.4)
United Kingdom	49 (5.0)	45 (5.7)	41 (6.0)
United States	35 (3.6)	30 (3.8)	29 (4.2)

New Zealand	30 (3.1)	26 (3.3)	23 (3.3)
Turkey	21 (2.1)	13 (1.7)	12 (1.7)
Other[b]	243 (24.7)	213 (25.5)	179 (26.0)

[a] Some have multiple roles.

[b] Other includes 63 countries: Philippines , India , Belgium , Germany , Italy , Netherlands , Poland , Hungary , Argentina , Egypt , Greece , Sweden , Thailand , France , China , Croatia , Lithuania , Russian Federation , Singapore , Bangladesh , Brazil , Chile , Israel , Nigeria , Syria , Uruguay , Colombia , Czech Republic , Ireland (Republic) , Japan , Kuwait , Malaysia , Mexico , Pakistan , Serbia , Slovakia , Slovenia , South Africa , Switzerland , Armenia , Austria , Belarus , Bolivia , Bosnia Herzegovina , Cameroon , Chad , Denmark , El Salvador , Indonesia , Iran , Korea South , Lebanon , Macedonia , Malawi , Morocco , Niger , Oman , Paraguay , Peru , Senegal , Venezuela , Vietnam , Yemen.

Box 1. Definitions of high rating outcome domains

Outcome domain	Definition
Vascular access problems	Problems with fistula, graft, or catheter required for dialysis e.g. access infections, bleeding, bruising, pain, discomfort, clotting
Death/Mortality	Number of people on hemodialysis who have died, risk of death, how long the patient will live
Cardiovascular disease	Disease of the heart and blood vessels e.g. heart attack, stroke, blockage of blood vessels
Dialysis adequacy	How well the dialysis cleans the blood, clearance, Kt/V
Fatigue/Energy	Feeling tired with no energy for weeks, for most of the time
Ability to travel	To go away for holiday, event, visiting family, work
Dialysis-free time	Time (hours/days) not doing dialysis

Table 3. Inclusion criteria for the core outcome set based on median, mean, proportion (7-9) and median rank scores

		Median score ≥ 8		**Mean score ≥ 7.5**		**Proportion % 7-9 score (critically important) ≥ 75%**		**Median rank (forced rank question)**	
		Patients/ caregivers	Health professionals	Patients/ caregivers	Health professionals	Patients/ caregivers	Health professionals	Patients/ caregivers	Health professionals
Core outcome set: Outcomes appear consistently (at least twice) across all measures for both groups	Vascular access problems	9.0	9.0	7.7	8.6	82	97	7.0	4.0
	Death/mortality	9.0	9.0	7.3	8.3	73	94	8.5	3.0
	Cardiovascular disease	8.0	9.0	7.6	8.4	77	95	7.0	3.0
	Dialysis adequacy	9.0	8.0	7.9	7.6	90	81	8.0	11.0
	Fatigue	8.0	8.0	7.6	7.6	82	86	8.0	11.0
Outcomes do not appear consistently (at least twice) across all measures for both groups	Ability to work	8.0	8.0	7.3	7.7	74	85	9.0	13.0
	Drop in blood pressure	8.0	8.0	6.9	7.9	68	89	11.0	9.0
	Ability to travel	8.5	7.0	7.7	6.8	75	57	10.0	17.0
	Dialysis-free time	8.0	7.0	7.6	7.1	74	67	9.0	15.0
	Infection/immunity	7.0	8.0	7.2	7.6	74	86	11.0	9.0
	Anaemia	8.0	7.0	7.4	7.3	77	76	11.0	10.0
	Mobility	8.0	7.0	7.3	7.2	76	71	11.0	14.0
	Blood pressure	8.0	7.0	7.4	7.3	76	81	10.0	9.0
	Washed out after dialysis	8.0	7.0	7.5	7.2	80	78	10.0	13.0
	Hospitalisation	7.0	8.0	6.6	7.6	65	86	13.0	8.0
	Impact on family/friends	7.0	7.0	7.1	7.2	66	72	11.0	15.0
	Depression	7.0	7.0	6.4	7.2	61	79	13.0	12.0
	Target weight	7.0	7.0	6.6	7.2	58	78	13.0	11.0
	Potassium	7.0	7.0	6.7	7.1	63	69	13.0	11.0
	Pain	7.0	7.0	6.9	7.2	66	75	13.0	12.0

Note: Grey shading indicates that the outcome met the criteria. I.e. the outcome rates consistently high, according to the inclusion criteria, for both patients/caregivers and health professionals. The outcome must appear at least twice for each of the criteria (median ≥ 8, mean ≥ 7.5, proportion ≥ 75%, and median rank < 10).