THE SONG HANDBOOK

FOR ESTABLISHING AND IMPLEMENTING CORE OUTCOMES IN CLINICAL TRIALS ACROSS THE SPECTRUM OF CHRONIC KIDNEY DISEASE

VERSION 1.0 | 1st JUNE 2017

SONG
STANDARDISED OUTCOMES IN NEPHROLOGY

Clinical trials that report important and relevant **outcomes** can help patients and their clinicians make decisions about treatment.

Contributors: SONG Executive Committee – Jonathan C Craig (Chair), Sally Crowe, John Gill, Tess Harris, Brenda Hemmelgarn, Braden Manns, Roberto Pecoits-Filho, Allison Tong, Peter Tugwell, Wim Van Biesen, Angela Wang, David C Wheeler, Wolfgang C Winkelmayer; Coordinating Committee – Yeoungjee Cho, Nicole Evangelidis, Talia Gutman, Camilla S Hanson, Lorraine Hamiwka, Martin Howell, Angela Ju, Amy Kelly, Karine E Manera, Richard McGee, Emma O'lone, Angelique F Ralph, Gaya Raman, Benedicte Sautenet, Daniel Sumpton, Armando Teixeira-Pinto, Andrea Viecelli. We also thank the many patients, caregivers, and health professionals involved in SONG for their contribution to the content in this Handbook.

Contact: SONG Initiative. Centre for Kidney Research, The Children's Hospital at Westmead, Westmead NSW, Sydney, Australia, 2145.

Suggested citation: SONG Initiative. The SONG Handbook. Version 1.0 June 2017, Sydney, Australia Available at songinitiative.org/reports-and-publications/

Executive Summary

The Standardised Outcomes in Nephrology (SONG) is an independent and global initiative that brings together patients, caregivers, and health professionals in partnership to establish core outcome domains and outcome measures across the spectrum of kidney disease for clinical trials and other forms of research.

A core outcome set is an agreed standardised set of outcomes that should be reported, as a minimum, in all trials within a specific area of health because they are critically important to all stakeholders. An outcome is something that can be measured, and can arise or change because of a health condition or treatment. Trials that report meaningful and relevant outcomes can help patients, families, and their clinicians make informed decisions about treatment based on outcomes that matter to them.

The SONG Handbook is for those who want to know about the SONG process, including individuals and organisations interested in developing and using core outcome sets within nephrology and in other areas of medicine. The Handbook describes the process for establishing, disseminating, and implementing core outcome domains and measures; and is living document that captures the methodological developments, knowledge, and the learning gained along the way.

SONG
STANDARDISED OUTCOMES IN NEPHROLOGY

Contents

1 | Background

1.1 What is the SONG Initiative?

The Standardised Outcomes in Nephrology (SONG) initiative aims to establish a set of core outcomes and outcome measures across the spectrum of kidney disease for clinical trials and other forms of research. The outcomes will be identified based on the shared priorities of patients, caregivers, clinicians, researchers, policy makers, and relevant stakeholders. This will help to ensure that researchers report outcomes that are meaningful and relevant to patients with kidney disease, their family, and their clinicians; and thus support decisions about treatment.

1.2 Problems with how outcomes are reported in trials

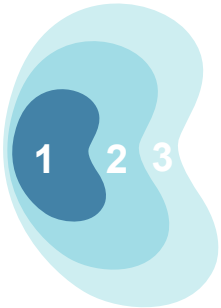
In clinical trials, treatments are tested by researchers to make sure they work and are safe. Researchers look at the effects those treatments have on patients and do this by measuring an "outcome". An outcome is something that can be measured, and can arise or change because of a health condition or treatment.

However, there are fundamental problems in the way outcomes are reported in trials, which are outlined in the following:

- Many outcomes reported in trials are not meaningful to patients, caregivers and health professionals for treatment decision-making. Surrogate or biochemical outcomes (e.g. potassium, calcium, phosphate in the blood) are frequently reported because they are easier to measure and are expected to be more responsive to the intervention[1]. However, most serum biomarkers have not been well validated and may not translate into health and quality of life outcomes that are meaningful to patients.
- The outcomes selected for trials, and the way they are measured and reported are inconsistent, which makes it difficult to compare the results across trials. In this way, studies cannot robustly inform decisions about treatment, and research efforts become inefficient[2].
- Bias in reporting outcomes occurs when outcomes are selectively reported when they favour the intervention, resulting in an overestimation of the effect of an intervention. Reporting bias also occurs when researchers do not report or change the outcome in the publication of the trial that was previously specified as a primary outcome in the protocol or in a trial registry. Patients may be at risk of harm if adverse events are not being reported[3].

1.3 The need for core outcomes

A core outcome set is an agreed standardised set of outcomes that should be measured and reported, as a minimum, in all clinical trials in specific areas of health or health care[4]. The core outcomes are of critical importance to all stakeholder groups i.e. patients/caregivers and health professionals. (Figure 1) Researchers can add other outcomes to the core outcome set.



1 CORE OUTCOMES
Critically important to all stakeholder groups. Report in all trials.

2 MIDDLE TIER
Critically important to some stakeholder groups. Report in some trials.

3 OUTER TIER
Important to some or all stakeholder groups. Consider for trials.

Figure 1a. Conceptual schema of a core outcome set (adapted from OMERACT)

Key points

- Core outcome domains are based on the shared priorities of patients, caregivers and health professionals that are critically important for decision making. Core outcomes are considered absolutely critical for all trials, as prioritised by patients, caregivers and health professionals.
- Researchers who conduct trials may add other outcomes based on different considerations (e.g. feasibility, responsiveness to the intervention).
- Core outcomes are not required to be used as primary outcomes (i.e. a core outcome does not have to be used as primary outcome to estimate the sample size necessary for an adequately powered study).

Researchers want to know if the specific intervention works. Patients/caregivers want to know if the intervention impacts on outcomes that matter to them.

1.4 Other initiatives to establish core outcomes

In recent years, there has been a proliferation of discipline-specific and global initiatives to develop core outcome sets for clinical trials. The Outcome Measures in Rheumatology (OMERACT) initiative was formed in 1992 and set the foundation for the development of core outcomes, specifically in rheumatology trials. With the engagement of patients, health care providers, and policy makers, OMERACT has improved the relevance of outcomes reported in rheumatology trials. More recently, the Core Outcome Measures in Effectiveness Trials (COMET) initiative was established to facilitate the development and collation of core outcome sets across all diseases internationally. COMET manages a database of initiatives and resources for developing core outcome sets, and they encourage core outcome developers to register their initiative on the COMET database. In 2016, The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) initiative collaborated with COMET to develop a guideline on how to select outcome measurement instruments[5] for core outcomes.

1.5 The SONG streams

Each SONG stream is based on a target population and this can defined by the CKD/treatment stage (i.e. chronic kidney disease Stage 1-5 [SONG-CKD], haemodialysis [SONG-HD], peritoneal dialysis [SONG-PD], kidney transplantation [SONG-Tx]), age (i.e. paediatrics [SONG-Kids]), or diagnosis (e.g. polycystic kidney disease [SONG-PKD]). Other sub-studies may be considered for specific types of interventions or patient groups. Proposals for SONG streams are considered and approved by the Executive Committee.

1.6 Outline of the SONG process

The SONG process is based on the validated methodology developed by OMERACT. The methodology for establishing core outcome domains involves: 1) systematic review to identify outcomes that have been reported in trials; 2) focus groups with nominal group technique involving patients and caregivers to identify, rank, and describe reasons for their choices; 3) stakeholder interviews to elicit perspectives on core outcomes; 4) international online Delphi survey to distil and generate a prioritised list of core outcome domains based on consensus; and 5) consensus workshop/s with stakeholders to review the core outcome domains and discuss implementation strategies. The development of core outcome measures follows a similar approach, and may involve validation studies.

The process is designed to capture the full range of outcomes regarded as important by all stakeholders, and also to prioritise the most critical outcomes that should populate all trials within the given CKD/treatment stage. For each stream, approximately 30 important outcomes will be identified and 3 to 5 will progress to the core outcome set for reasons of feasibility. Other trial-specific outcomes can also be reported by triallists.

The reporting of SONG studies will be guided by the Core Outcome Set-Standards for Reporting (COS-STAR) statement[4] and other relevant reporting guidelines (e.g. PRISMA for systematic reviews, COREQ for qualitative studies).

The process is underpinned by the values of: partnership, transparency, equity, trust, respect, evidence, and diversity.

1.7 Funding and support

SONG activities are supported by funding from government, philanthropic organisations, and professional societies.

1.8 The starting point – first meeting of the SONG Executive

The SONG Initiative was launched in November 2014 when the SONG Executive Committee met face-to-face for the first time in Philadelphia, United States, during the American Society of Nephrology Kidney Week. The Executive Committee decided to focus the first SONG stream on haemodialysis (SONG-HD).

2| Establishing core outcome domains

2.1 Introduction

For each SONG-stream (e.g. SONG-HD, SONG-Tx, SONG-Kids, SONG-PD, SONG-PKD, SONG-CKD), a set of consensus-based core outcome domains will be established. For feasibility, approximately 3 to 5 core outcome domains are identified for each stream. All other outcome domains identified are classified as middle or outer tier outcomes as shown in Figure 1.

2.2 Forming the Steering Group

A Steering Group is convened to develop core outcome domains within each of the streams of SONG. The Executive Committee appoints the Chair/Co-chairs of the Steering Group. The Chair will have strong expertise and experience in the specific CKD stage/treatment. In consultation with the Executive Committee, the Chair selects members based on their interest and expertise in the area. Individuals can self-nominate but final decisions will be made by Chair. The Steering Group should have international representation and consist of at least 10 members. Members will include:

- Clinicians with content expertise and experience in the outcome domains (multidisciplinary)
- Researchers (triallists)
- Patients/caregivers

All Steering Group members are listed on the SONG website.

2.3 Developing the protocol

The project coordinator develops the detailed protocol in consultation with the Steering Group and input from the coordinating committee. The protocol is published in a relevant peer-reviewed journal (e.g. *Trials* or specialty journal).

2.4 Systematic review

A systematic review is conducted to determine the range and heterogeneity of outcomes reported in randomised trials conducted in the target population of the respective SONG stream (e.g. haemodialysis, kidney transplant recipients, peritoneal dialysis). An assessment of outcome reporting bias may also be conducted if feasible and appropriate.

Selection criteria

Trials may be identified via electronic databases (Cochrane Kidney and Transplant Specialised Register, MEDLINE, Embase, PsycINFO, CINAHL), Clinicaltrials.gov, or from Cochrane Reviews. The trials should include more than 50% of the target population (e.g. patients on chronic haemodialysis [SONG-HD], kidney transplant recipients [SONG-Tx], patients on peritoneal dialysis [SONG-PD], children with chronic kidney disease [SONG-Kids], patients diagnosed with polycystic kidney disease [SONG-PKD], patients with chronic kidney disease [SONG-CKD]). The sampling frame (e.g. time frame) is decided upon by the Steering Group. It is recommended that at least 100 trials are included in the systematic review to generate sufficient data for credibility.

Data extraction

Characteristics of each trial are extracted and this may include (and is not limited to): publication year, study type, country, sample size, mean age of the participants, study duration, and type of intervention using a standardised data extraction template (available on request). All outcomes are extracted from every trial arm as reported in the trial. This includes the outcomes and the outcome measure (including outcome domain, measurement, metric, method of aggregation, and time points of measurement). Refer to section 6 for the SONG nomenclature of outcomes.

Data analysis

The outcomes are categorised into outcome domains. An outcome domain is a broad term that includes a set of specific outcomes/outcome measures (e.g. mortality, cardiovascular disease, pain). One investigator develops the list of the outcome domains, which is cross-checked by at least two other investigators until consensus is achieved. The outcome measures are grouped according to the final list of outcome domains, which are re-reviewed by the investigators. The outcomes are categorised into surrogate, clinical, and patient-reported outcomes as per the SONG nomenclature (see section 6). As outcome domains may span multiple categories, categorisation may be based on the largest proportion of outcome measures. Frequency of reporting is determined by calculating the number of trials that reported the outcome domains. Within each outcome domain, the specific outcome measures and time points of measurement are also analysed (Figure 2a). Further analysis (i.e. primary outcomes, subgroup analysis by time, trial characteristics etc, and outcome reporting bias) may also be done as is feasible and appropriate.

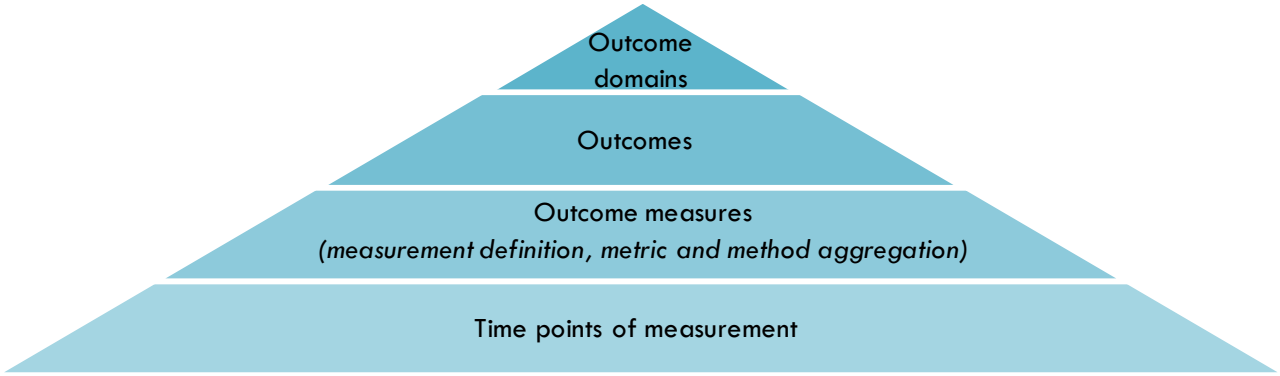


Figure 2a. Categorisation of outcome domains, outcomes, outcome measures, and time points

2.5 Focus groups with nominal group technique

Focus groups with nominal group technique are convened with patients/caregivers to identify and rank outcome domains that they consider to be important, and to discuss the reasons for their choices[6-9]. The nominal group technique is used to achieve consensus through generating, recording, discussing, and voting/ranking ideas. This approach minimises dominance of the discussion by individuals and fear of direct rejection of ideas and opinion. This technique has been used to prioritise outcomes[6-8].

Participant selection and recruitment

Patients/caregivers are recruited from participating sites. The location (national or international) of the sites as well as the language/s in which the focus/nominal groups are conducted depends on resources and collaborating investigators. Participants are purposively sampled to obtain a wide variation of demographic (i.e. age, gender, educational attainment, country) and clinical characteristics. The age of the participants is based on the specific stream. The participants should speak English (or in the language of the facilitator), able to give informed consent, and well enough to attend a focus group. Participants are reimbursed, on average, approximately $50 AUD to attend the focus group. Ethical approval is obtained from the Ethics Committee/Institutional Review Board of all participating sites.

Data collection

The focus/nominal groups are two hours in duration and are convened in a centrally located meeting room. Participants are seated around a table in a room with a whiteboard/flipchart and access to printing. Each session has three phases:

1) Discussion of the general experiences related to the specific CKD/treatment stage or diagnosis (or on a relevant topic)
2) Individual and group identification of outcome domains
3) Individual ranking of outcome domains

The facilitator for each group guides the discussion. At least one co-facilitator is also present to assist with administrative and logistical tasks (e.g. registration, printing), and to record non-verbal communication.

Facilitators are trained and observe at least one focus/nominal groups prior to moderating a group. Ideally, the same facilitator is present at all groups to ensure consistency of data collection across groups. Also, the knowledge and data gained from prior groups inform strategies for facilitating subsequent groups.

A detailed run sheet with the timing and questions is developed. Participants are asked to identify approximately 1 to 3 outcome domains they believe are important for trials (research). The ideas are listed on the board. The facilitator supplements the list with outcome domains reported in the systematic review (see 2.4). The list of outcome domains is discussed then printed. Participants individually rank the outcome domains in order of importance. If they are unable to rank the full list, participants are encouraged to rank at least the top 10.

All sessions are audiorecorded and transcribed verbatim.

The groups are convened until data saturation is achieved, defined as when no new outcomes or reasons for their choices are being identified in the subsequent groups.

The process may be adapted for the target population (i.e. appropriateness for SONG-Kids).

Data analysis

Quantitative data

The highest ranked outcome domain for each participant is assigned a value of 10, and the least important a value of 1. Outcome domains that were not ranked in the top 10 are given a value of 0. The individual rank scores for participants across all groups are used to determine mean rank score for the top 10 most important outcome domains from the combined list of outcome domains. As the number and type of outcome domains can vary, the mean rank score (ranging from 0 to 10) for each outcome is calculated based on the number of participants who ranked that outcome domain. The number of participants who ranked an outcome in the top 10 is also calculated. Sub analysis of mean rank scores may be performed by various characteristics (e.g. patients/caregivers, country, and gender). Statistical significance of differences may be assessed using a t-test with significance considered at $P<0.05$. As the rank scores for outcomes that are considered by a small number of groups may reflect the dynamic of the group, the mean values would provide a biased estimate of the relative importance of those outcome domains. Thus, a threshold for the number of groups may need to be determined for calculating the mean rank scores for the main results.

Qualitative data

The transcripts are entered into software for qualitative data management (e.g. HyperRESEARCH, NVivo). Using thematic analysis, the lead investigator reviews the transcripts line-by-line and inductively codes concepts that relate to the reasons for the participants' choices and ranking of outcomes. Similar concepts are grouped into themes. The preliminary results are discussed with multiple investigators to ensure that the analysis captures the full breadth and depth of the data collected.

2.6 Delphi survey

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[10] and has since been increasingly used as a valid approach to develop consensus-based core outcomes for clinical trials in various medical specialty areas[11].

Participant selection and recruitment

Inclusion criteria: Stakeholders including patients, caregivers/family members, nephrologists, surgeons, nurses, social workers, psychologists, dietitians, pharmacists (and other multidisciplinary clinicians relevant to the stream), policy makers, researchers and industry, with experience or interest in the specific CKD/treatment stage (HD, Tx, Kids, PD) are invited to join the Delphi Panel. The participants are aged over 18 years (except for SONG-Kids) and able to complete an online survey. The Delphi surveys are conducted in English language (other languages may be considered if feasible). All participants provide informed consent.

Selection strategy: Multiple strategies are used to recruit participants for the survey.

1. SONG database: an invitation is sent to the SONG database, which includes the names and emails of individuals who have registered to be involved in the SONG initiative. Currently, this includes approximately 3000 registrants.
2. Collaborating organisations: Relevant patient or professional organisations may help to reach out to participants via website posts, social media, and email circulation.
3. Recruitment sites: Patients can be directly identified and recruited from participating hospital/university institutions. This may require ethics approval.

Note: Target sample size: Previous Delphi surveys for core outcomes range from approximately 10 to 200. For each SONG-stream, the minimum target number of participants is 1000 with the aim of recruiting at least 50% patients/caregivers if feasible. To maximise reliability of the results, the target is to sustain a response of rate at least 70% across the three rounds.

Data collection

An outline of the Delphi process is shown in Figure 2a. However, this may be adapted within each stream as decided upon by the Steering Group.

Selection of outcome domains: The outcome domains are selected based on the prior phases (2.4 to 2.6). The list of outcomes is reviewed by the Steering Group and should be piloted with at least 10 patients/caregivers and 5 health professionals.

Note: The name and description of the outcome domains should be "framed" consistently to minimise bias responses. The reading age should be targeted to Grade 8 (or 12 years of age).

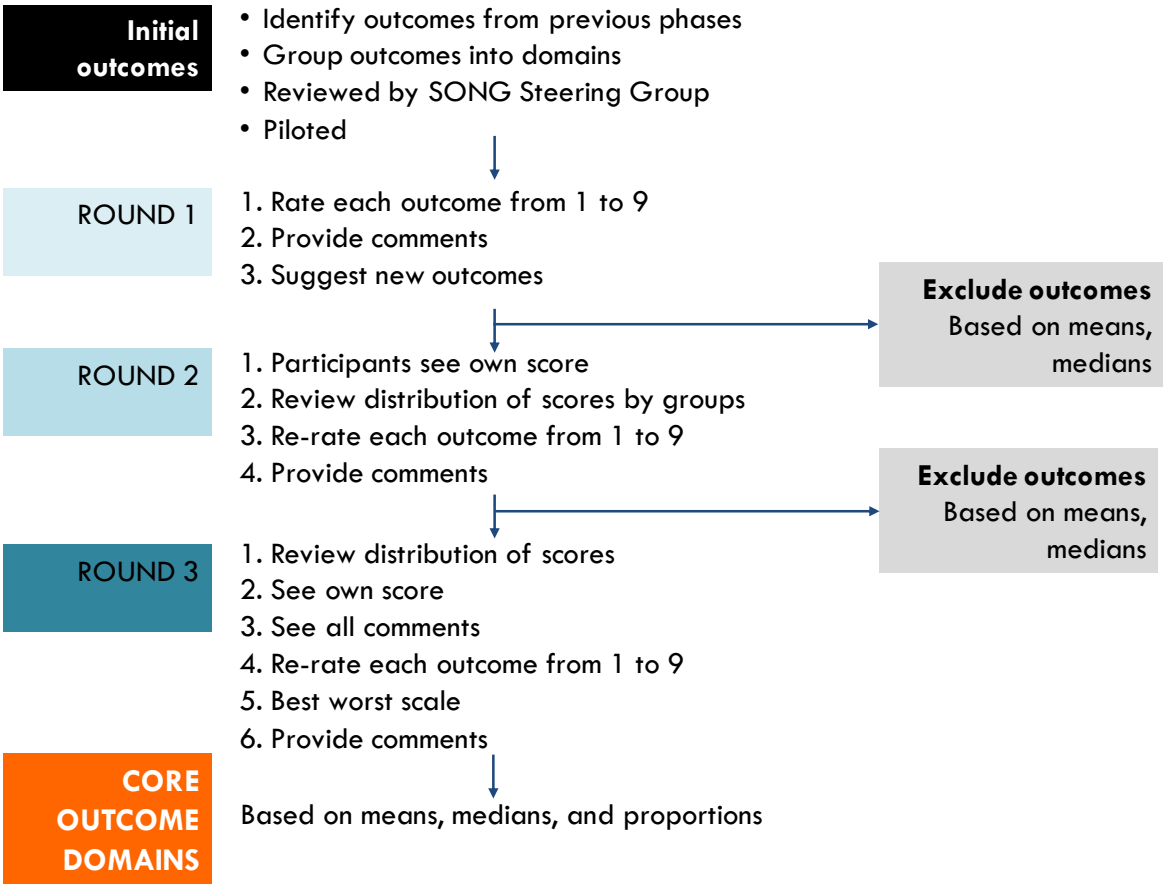


Figure 2b. Flowchart of the Delphi survey

To maximise participation, accuracy, and efficiency, the survey is administered online. The survey is also custom programmed as the individual responses are linked across rounds. The survey is conducted in

English. Translation of the Delphi survey into other languages depends on feasibility (i.e. resources, IT programming, timing of the surveys). Screenshots of the Delphi survey are provided in Appendix 2a. The outcome domains are randomised.

Round 1: Participants rate the importance of each outcome domain on a 9-point Likert scale using the GRADE scale (Rating 1 to 3 indicates "limited importance", 4 to 6 "important, but not critical," and 7 to 9 "critical importance")[12]. Comments can be provided in free-text responses. Participants are also asked to suggest new outcome domains. To distil the list of outcome domains to the most important based on consensus, outcomes with a mean and median of more than seven are taken through to Round 2. This threshold may need to be adjusted depending on the distribution of scores, and in consultation with the Steering Group. Any new outcome domains suggested by more than 10% of the participants are also taken through to Round 2.

Round 2: Participants see their own score from Round 1 and the distribution of the group responses on a column graph by patients/caregivers and health professionals, and by the weighted score across both groups. Participants can also see comments provided by patients/caregivers and health professionals in separate scroll down text boxes. After reflecting on these data, participants re-rate the importance of each outcome domain. Comments can be provided in free-text responses. Outcome domains that had a mean and median of more than 7, and had more than 50% rating the outcome domain 7 to 9, can be considered to be taken through to Round 3. Again, this threshold may need to be adjusted depending on the distribution of scores and in consultation with the Steering Group.

Round 3. Participants are provided with their own score, the group responses, and comments as per Round 2. Participants re-rate the final set of outcome domains.

Best Worst Scale: To quantify the relative importance of outcomes, participants are asked to complete a Best Worst Scale survey[13]. This has previously been completed in Round 3. However, it may be introduced in earlier rounds as decided upon by the Steering Group. Participants are presented with a maximum of six lists, each of which will contain a subset of approximately six outcome domains. Participants select the most important and the least important outcome domain from each list. To minimise survey burden, the best-worst scale survey uses a balanced incomplete block design[14] split into approximately four blocks randomly assigned to the participants. The BWS provides additional information on individual preferences and priorities and the influence of factors such as the individual perspective and experience, and cultural background.

Note: This process may need to be adapted (i.e. for age-appropriateness in SONG-Kids).

Data analysis

Quantitative

The mean, median, and proportion of participants who rated 7 to 9 for each outcome for all three rounds are calculated separately for patients/caregivers and health professionals. The mean difference in rating scores are assessed using the Wilcoxon signed rank test or a t-test with significance at $p<0.05$. Subgroup analysis may be performed for various characteristics (e.g. country, age, treatment).

For the best worst scale survey results, multinomial logistic regression models are used to calculate the relative importance score for each outcome domain normalised to the range of 0 (least important) to 10 (most important). This is also calculated separately for patients/caregivers and health professionals. The influence of factors (e.g. age) are also assessed.

Definition of consensus

Consensus on critically important outcomes is determined based on the mean, median, proportion of participants rating the outcome from 7-9 (critically important), and results from the Best Worst Scale survey. As the distribution of scores is unknown until after Round 3, it is not possible to provide an a priori definition of consensus to propose the set of 3 to 5 core outcome domains. The definition of consensus will also be discussed at the consensus workshop.

Qualitative

The qualitative data can be analysed using thematic analysis. The comments are entered verbatim into software for qualitative data management to facilitate cording. An investigator codes the comments and inductively identifies concepts focusing on the reasons for the ratings, changes in rating across rounds, and differences between patients/caregivers and health professionals. The preliminary analysis is discussed with multiple investigators to ensure the themes reflect the full range and depth of the data.

2.7 Consensus workshop/s

The SONG consensus workshop engages stakeholders in reviewing and discussing the proposed core outcome domains[15]. The workshop report will describe and summarise the perspectives of stakeholders on establishing and implementing the set of core outcome domains. This allows a better understanding of the potential challenges in establishing and implementing core outcomes, promotes acceptance, and informs strategies to optimise uptake and translation of the core outcome domains into clinical trials.

The consensus workshop/s are scheduled during major international conferences (e.g. American Society of Nephrology Kidney Week [SONG-HD], American Transplant Congress and the Congress of the International Transplantation Society [SONG-Tx], International Society of Peritoneal Dialysis Conference [SONG-PD]. The workshops are convened external to the conferences or as a satellite meeting depending on restrictions. The number of consensus workshops for each SONG stream depends on feasibility, location/timing of major conferences, and is decided upon by the Steering Group.

Participants and contributors

Patients, caregivers, and health professionals (including multidisciplinary clinicians, policy makers/regulators, industry) with experience and expertise relevant to the CKD treatment/stage or diagnosis addressed by the specific SONG stream are invited. To maximise potential for dissemination and implementation, key decision makers in professional societies, regulatory agencies, guideline organisations, trial networks, registries, and journal editors are invited to attend the workshop. Also, health professionals represent a wide range of different countries.

Approximately 70 participants attend the workshops with a target of at least one third being patients/caregivers. Invitees who are unable to attend are invited to be a contributor by providing feedback on the workshop program and draft report. All attendees and contributors are asked to sign consent to be a named investigator on publications arising from the workshop, and to be recorded and photographed for SONG-related publications.

Workshop program and process

All attendees and contributors receive a copy of the workshop program and materials one week prior to the workshop. During the workshop, the process and the preliminary results of the specific SONG stream is presented.

Attendees are allocated to breakout groups of approximately 10, which include a mix of patients/caregivers and health professionals. This can ensure a more dynamic and broader exchange of ideas. At least one member of the Steering Group (or Executive Committee) is present at each group to clarify and answer questions about the SONG initiative.

The facilitators and co-facilitators are provided with a detailed run sheet approximately one week prior to the workshop. A hard copy is given during a briefing session. The break out session broadly covers the following:

1. Welcome and introduction of each member
2. Feedback and clarification on the results (preliminary core outcome domains)
3. Implementation

Specific questions and prompts may be discussed in the breakout discussions as needed based on the preliminary results. All discussions are audio-taped and transcribed verbatim.

Synthesis of the workshop discussion

The lead workshop investigator reviews the transcripts line-by-line to inductively identify preliminary concepts. Similar concepts are grouped into themes. The transcripts are then imported into software to facilitate coding, and coded into the themes. The preliminary analysis is formatted into a draft workshop report, which is emailed to the workshop investigators (attendees and contributors) for feedback and comment. This ensures that the analysis reflects the diversity of opinion. Any additional feedback is integrated into the final report.

In the workshop report, selected quotations should be provided for each theme. Also, the key implications derived from the workshop are identified.

2.8 Establishing the set of core outcome domains

The proposed set of core outcome domains will be circulated in the draft workshop report and sent to all investigators and attendees for feedback and comment. Box 2a outlines criteria and considerations in proposing core outcome domains based on discussions at SONG consensus workshops[15].

Box 2a. Considerations and criteria for establishing core outcome domains

A core outcome domain should:

- Be critically important to the majority of all patient populations
- Have a clear, precise, and standardised definition that is understood by all stakeholders
- Be conceptualised by all stakeholder groups in a consistent way
- Be relevant over a longer time-frame (i.e. can be a short-term and long-term outcome)
- Be relevant across different countries and settings (including cultures, languages)
- Be single-attribute (i.e. does not include multiple outcome domains)
- Not be in direct conflict with another high-priority outcome
- Have broad relevance to a range of interventions for the target population
- Be feasibly applied in different types of trials (including pragmatic or registry trials)
- Be applicable in the context of assessing quality of care (e.g. quality indicator)
- Be considered to drive the research agenda, as well as to be reported in current trials

The set of core outcome domains should:

- Include a patient-reported outcome
- Include mortality as it is inherently fundamental to all other outcomes (refer to Appendix 2b)

The core outcome domains are posted on the SONG website. Please refer to section 5 for strategies to disseminate and implement the core outcome domains. To date, the SONG-HD (Figure 2c) and SONG-Tx (Figure 2d) have been established.

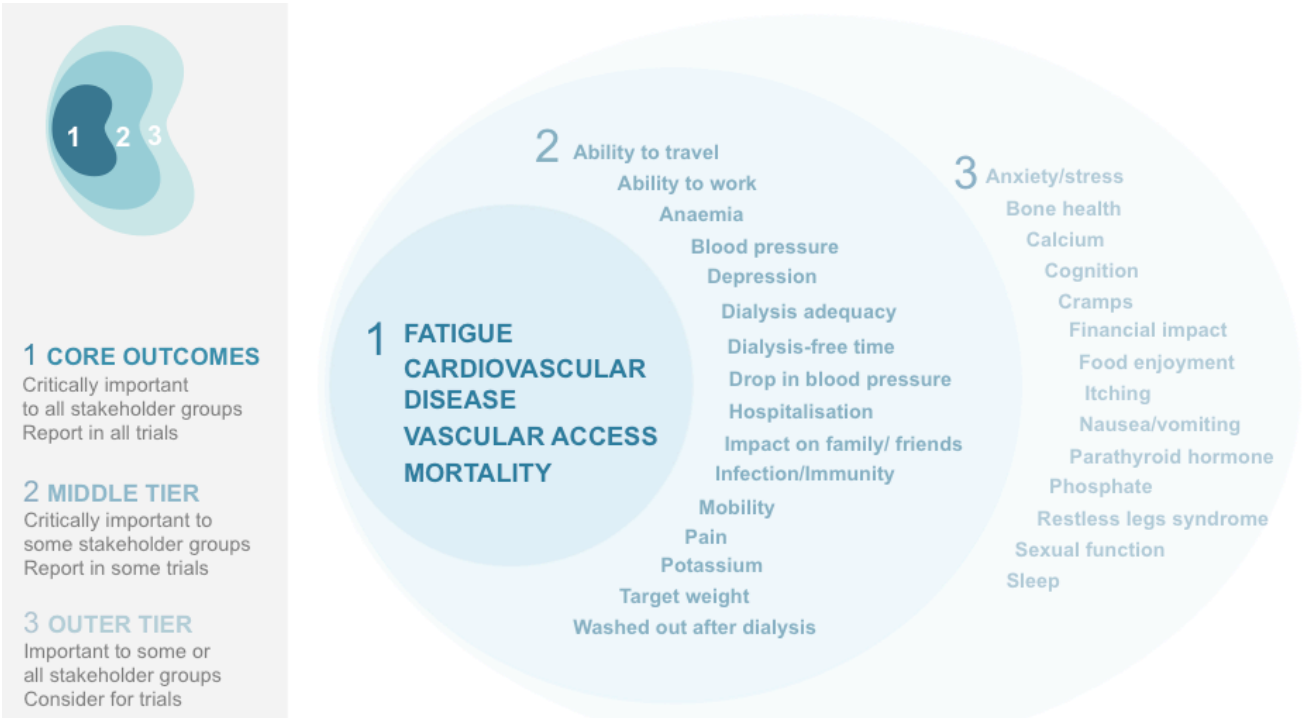


Figure 2c SONG-HD core outcome domains

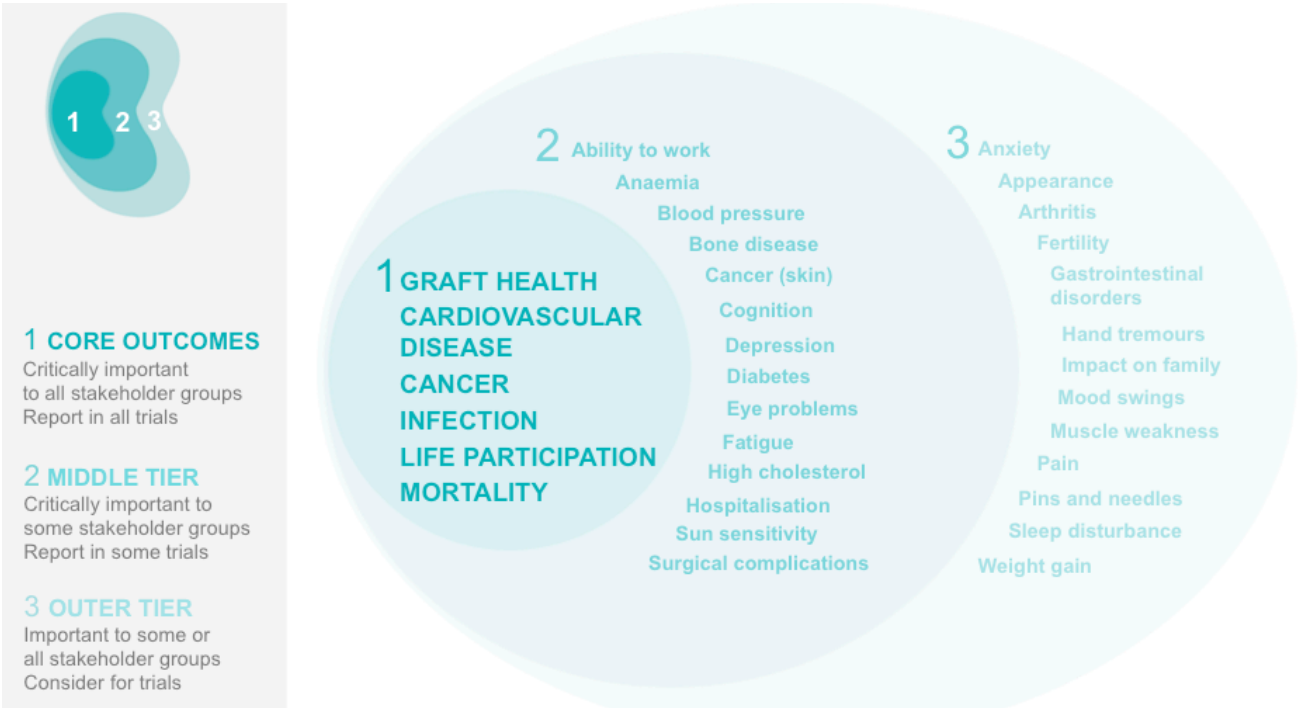


Figure 2d. SONG-Tx core outcome domains

3| Establishing core outcome measures

3.1 Introduction

A core outcome measure (or if appropriate, measures) is identified for each core outcome domain within a SONG stream (e.g. haemodialysis, kidney transplantation).

Why is it important to choose the "right" outcome measure for a core outcome domain?

An outcome measure is used to assess the effectiveness or safety of the intervention (e.g. treatment) being used in a trial. Choosing the "right" outcome measure can:

- help patients/clinicians measure the outcome domain in a *meaningful*, *appropriate,* and *easy* way as *accurately* as possible;
- guide the development and evaluation of interventions and patient care; and
- inform conversations and decision making about treatment plans.

The following provides an outline for developing a core outcome measure and may be adapted as appropriate.

3.2 Expert Working Group

The Expert Working Groups are responsible for establishing core outcome measures for each core outcome domain established within each stream of the SONG Initiative. The Executive Committee appoints the Chair/Co-chairs of the Expert Working Group. The Chair will have strong expertise and experience in relation to the core outcome domain. In consultation with the Executive Committee, the Chair selects members based on their interest and expertise in the core outcome domain or outcome measurement. Individuals can self-nominate but final decisions will be made by Chair. Each Expert Working Group should have international representation and consist of at least five members. Members will include:

- Clinicians with content expertise and experience in the outcome domain
- Researchers (triallists)
- Methodological experts in outcome measurement (patient-reported outcomes)
- Patients/caregivers

All Expert Working Group members are listed on the SONG website.

3.3 Developing the protocol

The detailed protocol for a specific core outcome measure is developed in consultation with the Expert Working Group and with input from the Coordinating Committee. The process may depend on the methods used and prior results in establishing the core outcome domains (section 2).

3.4 Systematic review

A systematic review is conducted to assess the range and heterogeneity of outcome measures used for a specific core outcome domain across trials. The systematic review may be required if, for example:

- the sampling frame is different to the systematic review of core outcome domains
- the review covers different aims, scope, or level of detail; and
- the review is used to assess measures for patient-reported outcomes (as this will provide more detail with regards to measurement properties).

The review may also include an assessment of outcome reporting bias. For clinical outcome measures, detailed analysis may be conducted to assess the use of composite outcomes, and competing risk.

Selection criteria

Trials may be identified via electronic databases as listed in section 2.4 (similar to those used for the systematic review of core outcome domains). The sampling frame (e.g. time frame) is decided upon by the

Expert Working Group. It is recommended that at least 100 trials are included in the systematic review to generate sufficient data for credibility.

Patient-reported outcome measures. For patient-reported outcomes, non-randomised studies may be included if too few measures have been reported in trials. The measures may be broad or specific to the dimension (e.g. fatigue, life participation), or designed for use within the target or general population.

Data extraction

The characteristics of each trial are extracted and this may include: publication year, country, sample size, mean age of the participants, study duration, study type, and type of intervention using a standardised data extraction template. All outcome measures are extracted from all trial arms as reported in the trial. This includes the outcome/dimension and the outcome measure (including measurement, metric, method of aggregation, and time points of measurement). Refer to section 6 for the SONG nomenclature of outcomes.

Patient-reported outcome measures. For systematic review of patient-reported outcomes, the general characteristics (e.g. number of items, duration of the survey, cost), and psychometric properties are extracted.

Data analysis

Clinical outcome measures. Please refer to section 2.4.

Patient-reported outcome measures. All items related to the outcome domains (e.g. fatigue, life participation) are extracted from the measures. The full list of items is classified into dimensions. These dimensions are cross checked by at least one other reviewer and discussed until consensus is achieved. The frequency of each dimension across all the measures is calculated. The COnsensus-based Standards for the selection of health Measurement Instruments-Core Outcome Measures in Effectiveness Trials (COSMIN-COMET) guidelines[5] is used to evaluate the psychometric properties (content validity, reliability, responsiveness, internal consistency, structural validity, measurement error, criterion validity, and cross cultural validity) of each measure. A hand search for studies reporting psychometric data in the target population is also conducted. Studies of the initial development of the measures may be used to provide additional psychometric data.

3.5 Stakeholder survey

An international online survey is conducted to identify the most critically important outcome (or dimensions) within an outcome domain. Table 3a provides some examples of outcomes.

Table 3a. Examples of outcomes in an outcome domain

Outcome domain	Outcomes/dimensions*
SONG-HD Vascular access	• Function • Infection • Maturation • Bleeding
SONG-HD Fatigue	• Impact of fatigue on life participation • Ability to think clearly • Limb/muscle weakness
SONG-HD Cardiovascular disease	• Myocardial Infarction • Stroke • Heart failure • Peripheral vascular disease
SONG-Tx Graft health	• Graft loss • Graft function

*The term dimension may be used for patient-reported outcome measures

Participant selection and recruitment

Inclusion criteria: As per the Delphi survey (section 2.6), stakeholders including patients, caregivers/family members, nephrologists, surgeons, nurses, social workers, psychologists, dietitians, pharmacists (and other multidisciplinary clinicians relevant to the stream), policy makers, researchers and industry, with experience or interest in the specific CKD/treatment stage (HD, Tx, Kids, PD) can participate in the survey. The survey is conducted in English and at least one other language as determined by the Expert Working Group. All participants provide informed consent.

Selection strategy: As per the Delphi survey (section 2.6); multiple strategies are used to recruit participants for the survey.

1. SONG database: an invitation is sent to the SONG database, which includes the names and emails of individuals who have registered to be involved in the SONG initiative. Currently, this includes over 3000 participants.
2. Collaborating organisations: Relevant patient or professional organisations may help to reach out to participants via website posts, social media, and email circulation.
3. Recruitment sites: Patients can be directly identified and recruited from participating hospital/university institutions. This may require ethics approval.

Note: The target sample size is minimum 500 participants with at least 50% being patients/caregivers. For non-English surveys, the target sample size is 100 participants.

Data collection

Selection of outcomes: The outcomes/dimensions are selected based on the prior phases of establishing core outcome measures (3.4 – 3.5), and may also be informed by the findings in section 2 (i.e. data from the focus groups, Delphi survey comments), other literature, and with input from the Expert Working Group. Outcomes that were previously included in the Delphi survey (e.g. for the outcome domain vascular access, infection and hospitalisation was identified as an outcome but had been included in the Delphi survey), may be included based on discussion with the Expert Working Group. Inclusion may be justified based on the following:

- The Delphi comments (e.g. patients rated vascular access of critical importance because of infection or hospitalisation).
- A clinical or patient-reported outcome that has been reported frequently in trials.
- Consensus among the Expert Working Group that the outcome would be relevant, important, and potentially critical for decision-making.

The survey is piloted with a least 10 patients/caregivers and 5 health professionals. The outcomes are randomised in the survey.

Note: The name and definitions of the outcomes should not "bias" responses. This is considered in framing the outcome and describing consequences of the outcome.

Survey design: To maximise participation, accuracy, and efficiency, the survey is administered online. However, a paper-based survey may be considered if this is the only mode of administration that is feasible for collecting responses.

GRADE rating: Participants rate the importance of each outcome on a 9-point Likert scale using the GRADE scale (Rating 1 to 3 indicates "limited importance", 4 to 6 "important, but not critical," and 7 to 9 "critical importance")[12]. Comments can be provided in free-text boxes. Participants can suggest new outcomes and rate their suggestions.

Best worst scale survey: After rating the outcomes, participants complete a best worst scale survey to determine the relative importance outcomes. As detailed in section 2.6, participants are presented with a maximum of six lists, each of which will contain a subset of approximately six outcomes. Participants select the most important and the least important outcome from each list. The best-worst scale survey uses a balanced incomplete block design split into approximately four blocks randomly assigned to the participants.

Data analysis

Quantitative

The mean, median, and proportion of participants who rated 7 to 9 for each outcome for all three rounds are calculated separately for patients/caregivers and health professionals. The mean difference in rating scores are assessed using the Wilcoxon signed rank test or a t-test with significance at $p<0.05$. Subgroup analysis may be performed for various characteristics (e.g. country, age, treatment).

For the best worst scale survey results, multinomial logistic regression models are used to calculate the relative importance score for each outcome domain normalised to the range of 0 (least important) to 10 (most important). This is also calculated separately for patients/caregivers and health professionals. The influence of factors (e.g. age) are also assessed.

Qualitative

Depending on the content and volume of the qualitative data, a descriptive summary or a thematic analysis can be used to analyse and present the data.

3.6 Consensus workshop/s

The general objectives of the consensus workshop are to: provide an overview of the SONG initiative (including the specific stream); understand and discuss how to choose core outcome measures for a core outcome domain; and to suggest strategies for implementation.

Stakeholder input will include perspectives on (but not limited to):

- The absolute and relative importance of the top prioritised outcomes based on the international survey (i.e. support or confirm the importance of the outcome)
- Conceptualisation and clarification of the definition of the outcome
- Impact of the outcome
- Suggestions of outcome measures, feedback on proposed measures (where possible) and this may include feasibility, challenges of using and interpreting the measure

To maximise participation from different countries, the consensus workshop/s are scheduled during major international conferences. The workshops are convened external to the conferences or as a satellite meeting depending on restrictions. The number of consensus workshops for each SONG stream depends on feasibility, location/timing of major conferences, and is decided upon by the Steering Group.

Participants and contributors

Patients, caregivers, and health professionals (including multidisciplinary clinicians, policy makers/regulators, industry) with experience and expertise relevant to the CKD treatment/stage or diagnosis addressed by the specific SONG stream and core outcome domain are invited. To maximise potential for dissemination and implementation, key decision makers in professional societies, regulatory agencies, guideline organisations, trial networks, registries, and journal editors are invited to attend the workshop. Also, health professionals represent a wide range of different countries.

Approximately 60 participants attend the workshops with a target of at least one third being patients/caregivers. Invitees who are unable to attend are invited to be a contributor by providing feedback on the workshop program and draft report. All attendees and contributors are asked to sign consent to be a named investigator on publications arising from the workshop, and to be recorded and photographed for SONG-related publications.

Workshop program and process

All attendees and contributors receive a copy of the workshop program and materials one week prior to the workshop. The materials will include a background, outline of the methods, preliminary results, and proposed core outcome measures (refer to 3.7). During the workshop, the process and the preliminary results of the specific SONG stream is presented.

Attendees are allocated to break out groups of approximately 10, which includes a mix of patients/caregivers and health professionals. This can ensure a more dynamic and broader exchange of

ideas. At least one member of the Expert Working Group is present at each group to clarify and answer questions about the SONG initiative, or technical and content questions related to the outcomes.

The facilitators and co-facilitators are provided with a detailed run sheet approximately one week prior to the workshop. A hard copy is given during a briefing session. The break out session broadly covers the following:

1. Welcome and introduction of each member
2. Feedback and clarification on the results and the proposed core outcome measures (if available)
3. Implementation

Specific questions and prompts may be discussed in the breakout discussions as needed based on the preliminary results. All discussions are audio-taped and transcribed verbatim.

Synthesis of the workshop discussion

The lead workshop investigator reviews the transcripts line-by-line to inductively identify preliminary concepts. The concepts should be largely specific to core outcome domains and the proposed core outcome measures. Similar concepts are grouped into themes. The transcripts are then imported into software to facilitate coding, and coded into the themes. The preliminary analysis is formatted into a draft workshop report, which is emailed to the workshop investigators (attendees and contributors) for feedback and comment. This ensures that the analysis reflects the diversity of opinion. Any additional feedback is integrated into the final report.

In the workshop report, selected quotations should be provided for each theme. Also, the key implications derived from the workshop are identified.

3.7 Proposing the core outcome measures

The project coordinator works with the Expert Working Group to propose a core outcome measure. This may be done prior to the consensus workshop (3.6). The project coordinator and Expert Working Group review the proposed core outcome measure against the COSMIN-COMET Framework[5] (Table 3a, 3b). The Omeract Filter 2.0 (Box 2a) and examples of frameworks for developing an outcome measure (Appendix 3a) can also be considered. Selection should also be in consideration of stakeholder input from the previous phases (e.g. Delphi survey, consensus workshop).

Note: The proposed core outcome measure/s is explicitly justified against these criteria (or indicated as not applicable or unable to be assessed due to lack of evidence).

Table 3b. Measurement properties adapted from COSMIN-COMET 2016

Property	Definition
Content validity	The degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured.
Reliability	The degree to which the measurement is free from measurement error. I.e. whether measuring the outcome provides similar or the same result on multiple occasions under consistent conditions.
Responsiveness	The ability of a measurement instrument to detect change over time in the construct to be measured.
Internal consistency	The degree of interrelatedness among the items i.e. how closely different measures that assess the same outcome are related.
Structural validity	The degree to which the scores of a measurement instrument are an adequate reflection of the dimensionality of the construct to be measured.
Measurement error	The systematic and random error of a patient's score that is not attributed to true changes in the construct to be measured i.e. how much of the changes in measurement results are not due to true changes in the outcome but related to problems with accurate measurement.
Hypothesis testing	The degree to which the scores of a measurement instrument are consistent with

	hypotheses based on the assumption that the measurement instrument validly measures the construct to be measured.
Criterion validity	Whether the outcome measure is a good reflection of the "gold standard" or best instrument available for that measurement (provided that there is one).
Cross cultural validity	The degree to which the performance of the items on a translated or culturally adapted measurement instrument is an adequate reflection of the performance of the items of the original version of the measurement instrument i.e. whether an outcome measure can be used across cultures and reproduce the same results.

Table 3c. Feasibility aspects adapted from COSMIN-COMET 2016

Feasibility aspect	Definition and considerations
Patients' comprehensibility	The patient can understand it.
Interpretability	We can interpret the results.
Ease of administration	It is easy to administer.
Length of the outcome measurement instrument	It is not too long.
Completion time	The time it takes to complete.
Patient's mental ability level	The patient has the mental capacity to do it.
Ease of standardisation	It is easy to standardise.
Clinician's comprehensibility	The clinician can understand it.
Type of outcome measurement instrument	The type of measure.
Cost of an outcome measurement instrument	The cost to obtain and administer it.
Required equipment	The equipment needed.
Type of administration	How the measure is completed.
Availability in different settings	It is available in different settings.
Copyright	Whether it requires a license or permission to use.
Patients' physical ability level	The patient's physical ability to do it.
Regulatory agency's requirement for approval	Whether it requires regulatory approval.
Ease of score calculation	How easy it is to assess the results.
Invasiveness*	Inconvenience of doing it.

*Added to the original table.

Box 3a. The OMERACT Filter 2.1

The OMERACT Filter 2.1[16] assesses:

1. **Truth** (face, content, construct validity);
2. **Feasibility** (respondent burden, financial cost, interpretability of results, access, translation)
3. **Discrimination** (reliability and internal consistency, responsiveness [within-group discrimination], use in trials [between-group discrimination], thresholds of meaning); and

Key points from the OMERACT Handbook:

- A simple report of count of deaths is mandatory, and further exempt from Filter requirements. However any instrument assessing a more detailed specification of a Domain within the Area of Death (for example death from a specific cause) would need to pass Filter 2.0 and an instrument in that Domain would need to satisfy all the requirements.
- Where only partially validated instruments are identified for the setting of the Domain, or where no instruments are available in a Domain, instruments will need to be further validated respectively

developed and their applicability documented. Similarly, for any contextual factor or adverse event declared to be core, at least one applicable measurement instrument must be identified or developed.

- For specific measurement technique (e.g. MRI, ultrasound), focus closely on Discrimination, more specifically its potential to predict changes in other biomarkers or outcome, and feasibility. For biomarkers, often a clear understanding of the pathophysiology underlying the phenomenon may not be fully available. To meet the Truth part of the Filter the pathophysiologic documentation should be updated as insights evolve. In addition, technical aspects of measurement (e.g. detailed procedure, inter-assay variation) often need to be addressed before moving on to Filter validation.
- Assessment of applicability of instruments can be carried out in two phases: an initial screen to quickly determine the most likely candidates (and eliminate those extremely unlikely to prove adequate); and then a full evaluation, including an assessment of the approach to the instrument's development (for example, was a questionnaires developed through item response theory or through classical test theory and were patients included in its development, or how closely is a biomarker linked to the underlying pathology) and a thorough risk of bias assessment to discern if there are any fatal flaws in the studies documenting applicability of an instrument.

3.8 Validation of core outcome measures

The proposed core outcome measure must be validated before it can be recommended and endorsed for use. The criteria and process for validation will be outlined separate for patient-reported outcome measures and clinical outcome measures.

Patient-reported outcome measure:

Figure 3a (adapted from Rothrock et al[17]) outlines the process of validating a core patient-reported outcome measure.

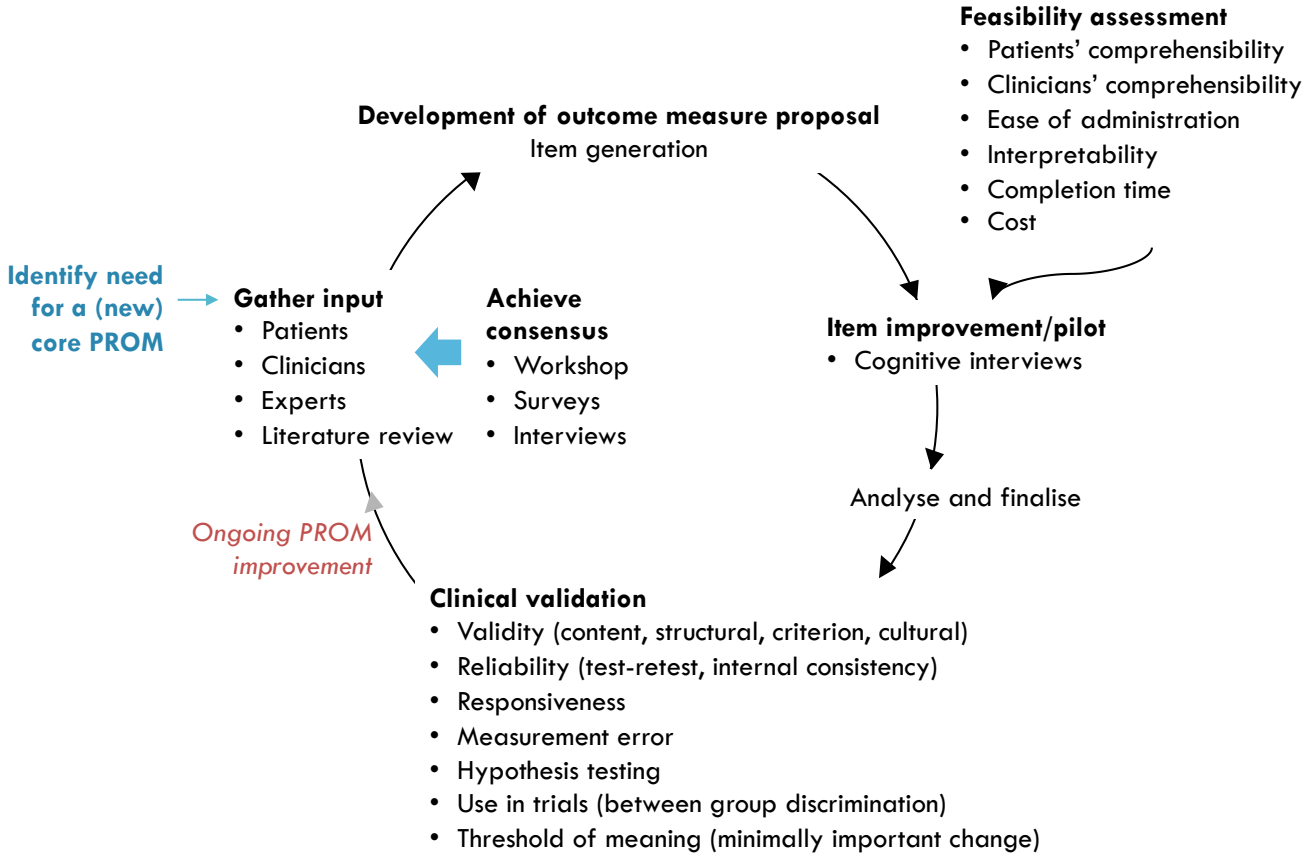


Figure 3a. Outline of the process for validating a core patient-reported outcome measure

Each core outcome set will include at least one patient-reported outcome, thus establishing the need for a core patient reported outcome measure (PROM) (e.g. to measure fatigue, life participation, pain). Stakeholder input and clinical validation data of an existing or a modified measure might be sufficient to choose an existing PROM to include in the core outcome set. A systematic review of existing outcome measures used in RCTs and/or non-RCTs and observational studies to assess the outcome (e.g. fatigue, life participation, pain) in the target population is conducted (Refer to 3.4). This will summarise the range and frequency of measures used, and currently available evidence on the psychometric properties of existing measures. This will also inform the development or adaptation of measures if required.

To generate relevant items (questions) for the core PROM, input is sought from all stakeholder groups including patients/caregivers and health professionals. Health professionals include clinicians and researchers with experience in the construct of interest (i.e. dimension being measured), and have expertise in psychometrics. Through discussion, the Expert Working Group and other stakeholders involved reach a consensus on how the construct should be assessed. This involves establishing a clear definition and scope of the construct, as well as developing a conceptual model where appropriate. There are various methods to reach consensus, including consensus workshops and the Delphi technique[18].

The initial measure developed from the items generated with input from stakeholders can be improved through piloting it in the patient population of interest. Feasibility aspects (patients' and clinicians' comprehensibility; ease of administration; interpretability; completion time; cost) are assessed in relation to the patient population and context of administration[5]. Cognitive interviewing is a useful tool to identify problematic questions in terms of comprehensibility as it involves asking patients to verbalise their thoughts whilst completing the measure[19,20]. Pilot study results are analysed to identify any floor/ceiling effects, unusual amount of variability and reasons for the anomalies.

The finalised version of the core PROM measure is then put through initial validation work. According to the COSMIN-COMET guidelines on how to choose an outcome measure, a range of measurement properties need to be examined: content, structural, criterion and cross-cultural validity, reliability, internal consistency, responsiveness, measurement error and hypothesis testing[5]. Additional properties include convergent validity (correlation of scores from existing measures that assess the same construct or similar concept) and divergent validity (lack of correlation with existing measures known to assess different construct[21]). To ensure meaningful application in trials, it is also necessary to define the methods of aggregation and meaningful thresholds, which requires that a minimally important clinical change (MICD) is known. Should an MICD not be known or available, then an appropriate strategy for development should be proposed[22]. Validation is a process that is built and strengthened over time, with more studies replicating or adding new evidence for a case that the measure is appropriate to assess a specific construct in the population of interest[17]. For all patient-reported outcome measures endorsed by SONG for the core outcome sets, SONG will take a pragmatic approach in achieving a balance between conducting necessary validation work and endorsing an outcome measure within a reasonable timeframe.

Clinical outcome measure:

Informed by the systematic review (refer to 3.4) and stakeholder input from previous phases (survey and workshop), members of the Expert Working Group conduct an independent assessment of the proposed outcome measures (i.e. definition, measurement, metric and method of aggregation) by judging the fulfilment of relevant measurement properties and feasibility aspects. The focus on individual properties may differ depending on the type of measure proposed. Table 3d includes properties considered for test method validation for patient reported outcomes, criteria established by the FDA[23] and medical device companies[21,24], and items from the COSMIN-COMET 2016 framework[5].

Based on consensus of the WorkingG members and Steering Committee, one to two core outcome measures will be selected for each core outcome. If a consensus is achieved on a measure that has been widely used (as determined by the systematic review, see section 3.4) then measurement properties such as reliability and validity of the selected core outcome measure can be assessed by post-hoc analyses of existing trials. If a new or partially altered outcome measure is proposed for which there is no gold standard or comparable outcome measure in the literature, the proposed core outcome measure is validated in a prospective clinical trial in the appropriate target population. Measurement properties such as internal consistency (Cronbach's alpha), construct validity (correlation coefficient), test-retest reliability

(intraclass correlation coefficient), responsiveness and feasibility aspects such as cost and ease of administration are examined[5]. As for PROMs, consideration of use of the measure within trials is also required including methods of aggregation and defining meaningful thresholds.

Table 3a. Appraisal of measurement properties and feasibility aspects of proposed core outcome measures

Proposed measurement definition		
Proposed specific metric		
Proposed method of aggregation		
Criteria	**Meets criteria***	**Comments (references)**
Measurement properties		
Content validity		
Reliability		
Responsiveness		
Internal consistency		
Structural validity		
Measurement error		
Minimally important clinical difference		
Hypothesis testing		
Criterion validity		
Cross cultural validity		
Feasibility aspects		
Patients' comprehensibility		
Interpretability		
Ease of administration		
Completion time		
Ease of standardization		
Clinicians' comprehensibility		
Cost		
Required equipment		
Type of administration		
Copyright		
Patients' mental ability level		
Patients' physical ability level		
Regulatory agency's requirement for approval		

*Indicate 1=meets criteria, 0=does not meet criteria, NA=not applicable, Unclear=unable to be assessed due to lack of evidence (this may be used to score)

3.9 Establishing the core outcome measure

The proposed core outcome measure is circulated in the draft workshop report and sent to all investigators and attendees, and the SONG network, for feedback and comment. Subsequently, the validated core outcome measure will be endorsed for use.

4| Stakeholder engagement

4.1 Introduction

The SONG initiative engages a broad range of stakeholders in establishing and implementing core outcomes across the spectrum of chronic kidney disease; to ensure that the outcomes reported in trials are directly relevant for treatment decision-making. Stakeholders are involved in all phases of SONG and at multiple decision-making levels. Efforts are made to ensure that stakeholders are empowered to contribute in a meaningful, equitable, and transparent way.

4.2 Who are the stakeholders?

Stakeholders (including stakeholder organisations) are individuals and groups with an interest in core outcomes for trials in chronic kidney disease and include:

- Patients
- Caregivers (family members, friend, or others who provide care for the patient in an unpaid capacity)
- Physicians (e.g. nephrologists, surgeons, psychiatrists)
- Nurses and allied health professionals (e.g. psychologists, social worker, dietitian)
- Policy makers
- Researchers (including clinical triallists)
- Industry

4.3 What is the purpose of stakeholder engagement in SONG?

The objective is to engage stakeholders (particularly patients/caregivers) in all stages of the SONG process (Figure 4a). Involving stakeholders early and throughout the process can also help to facilitate better acceptance and uptake of core outcomes. Stakeholders can provide perspectives and input on (but not limited to):

- The absolute and relative importance of outcome domains and outcomes (i.e. confirm the importance of the outcome).
- The definition of the outcome domain and outcomes.
- Impact of the outcome domain and outcome, especially from those living with CKD and their caregivers.
- Proposed measures (where possible) and this may include feasibility, challenges of using and interpreting the measure.
- Strategies for ensuring that core outcomes are used.

Figure 4a. Stage of involvement

4.4 How are stakeholders involved?

Stakeholders contribute across all stages of SONG in the following ways:

- As members on the Executive Committee, Steering Group, Expert Working Groups (at least 1-2 members should be patients/caregivers).
- Participating in focus group studies with nominal group technique to identify, rank, and discuss outcomes (patients/caregivers).
- Participating in stakeholder interviews.
- Participating in surveys (Delphi survey, outcomes surveys).
- Attending SONG Consensus Workshops.

4.5 Strategies for engaging and recruiting stakeholders

The SONG database was set up to record the names and email addresses of individuals who have been involved in any aspect the initiative (i.e. as committee members, participants) and those who registered their details on the website. The SONG website includes a webpage that invites stakeholders to join the SONG initiative http://songinitiative.org/get-involved/.

Multiple strategies are used to engage stakeholders.

- Sending invitations and information about opportunities to be involved to the SONG Initiative network to professional societies and patient organisations.
- Direct recruitment through hospital organisations – this has been the most effective way of reaching out to patients and caregivers to participate in surveys and workshops. Investigators (namely clinicians who know the patients), invite patients/caregivers to participate or obtain permission to submit their contact details for SONG-related research projects and activities. For surveys, updates of the demographic characteristics are sent periodically to investigators to target recruitment in countries with low response. If patients/caregivers are recruited through hospital institutions and are involved as research participants, ethics approval is likely to be required.
- Posting on website and social media platforms (e.g. twitter, Facebook).
- Disseminating invitations/website link through presentations.

4.6 Partnering with patients and caregivers

Patients and caregivers bring valuable experiential knowledge as they have firsthand insights on living with the disease and treatment, and how this impacts their well-being. Also, given the mismatch between patients and health professional priorities, working in partnership can help to ensure that patients' priorities are reflected in the core outcome set.

> *"It sounds like they took a cold medical approach where here it's more personal, there's shared experience, there's more heart that's been put onto those outcomes [we have prioritised]."* – Mother of child with CKD

In the SONG-HD process, patients and caregivers identified and gave high priority to outcomes that are rarely, if not never reported in trials including: dialysis free time, ability to work, ability to travel, and impact on family and friends[7,25]. Also, the SONG-HD Delphi Survey demonstrated striking discrepancies between patients/caregivers and health professionals in terms of how they defined and prioritised outcomes. For example, patients conflated dialysis adequacy with "feeling well" whilst health professionals defined the outcome in terms of urea kinetics, Kt/V, and fluid[15]. Patients/caregiver gave higher priority to ability to travel and dialysis-free time compared with health professionals who rated outcomes such as mortality and hospitalisation higher than patients/caregivers[25].

Table 4a. outline how patients/caregivers are engaged across the stages of SONG.

Table 4a. Patient/caregiver involvement in the SONG process

Stage	Mode and type of input from patients/caregivers
What to do	• Executive committee, Steering Group, Expert Working Group - Discussing and identifying priority streams
How to do it	• Executive committee, Steering Group, Expert Working Group – Providing input on the protocol
	• Piloting – Providing feedback on surveys
	• Recruitment – Suggesting strategies for recruiting participants
Doing it	• Participating in focus groups to identify outcomes • Disseminating invitations to participate in the Delphi survey • Participating in consensus workshops and contributing opinions
Reviewing it	• Providing feedback on draft reports (surveys, workshop reports)

	•	Consensus workshop – providing feedback on preliminary results
Disseminating	•	Journal publications – Contributing as co-authors on manuscripts
	•	Conference presentations – Presenting findings at conferences of patient events
Implementing	•	Consensus workshops – Identifying strategies for implementation
Evaluating	•	Feedback on including core outcomes in trials

Maximising meaningful involvement of patients/caregivers may require additional resources to provide education and training, and reimbursement for participating (at face-to-face discussions including focus groups and consensus workshops).

Involving patients/caregivers across all stages of SONG provides an opportunity to change how trials select outcomes, which will ultimately result in higher-level impact on research in chronic kidney disease more broadly. The core outcome set, with input from patients/caregivers, will help to:

- Inform what outcomes trials should measure and report, which can ultimately facilitate the use and implementation of trial evidence in practice and policy; and potentially
- Drive the research agenda to focus on outcomes that are critically important.

4.6 International representation

The focus groups, Delphi survey and workshops involve multinational sites when possible. Whilst the primary language is English, efforts are made to conduct studies in other languages though this depends on the resources, feasibility, and target patient population.

It would not be feasible to achieve complete representation from all countries. Potential bias may be acknowledged, however the SONG process involves a relatively large number of participants from a range of countries and relevant stakeholder groups.

5| Dissemination and implementation

5.1 Introduction

Multiple strategies are needed to maximise dissemination and implementation of the core outcome domains and measures in trials. COMET recommends that developers register the initiative on the COMET database and to "consider engagement with the relevant Cochrane Review Groups, clinical guideline developers, research funders, journal editors, regulators such as research ethics committees, and trial registries." Of note, the NIHR Health Technology Assessment funding body in the UK has recently added the following statement to its application form: "Details should include justification of the use of outcome measures where a legitimate choice exists between alternatives."

Table 5a outlines strategies for implementing SONG core outcome domains and measures.

Table 5a. Outline of strategies and mechanism for implementing SONG core outcome domains and measures

Mechanism	Suggested strategies and actions
General	• Include a link to the SONG Website
Journals	• Solicit an editorial or commentary regarding the SONG outcome domains/measures • Include mention of the SONG outcome domains in the authorship policy (for publication of clinical trials)
Guidelines	• Include mention of the SONG core outcome domains/measures in resources provided for developing guidelines and other guideline-related publication and activities
Trial networks	• Advise investigators to include SONG core outcome domains outcomes/measures in protocols and the scientific review of trials
Trial registries	• Include a reference to core outcomes
Professional societies	• Send a notification to members about the SONG core outcome domains/measures (via email, newsletter)
Research organisations	• Advise to include SONG core outcome domains/measures at protocol review (e.g. Cochrane) • Add SONG (all streams, projects) to the COMET database
Registries	• Include SONG core outcome domains/measures in data collection forms • Include mention of SONG core outcome domains/measures at registry events and other related initiatives • Send a notification to relevant individuals/groups (e.g. working groups) about the SONG core outcome domains/measures
Funders	• Reference core outcome domains/measures in guidance documents for application and peer review • Targeted calls for funding that focus on core outcomes
Policy/ regulators	• Minimum requirements for data collection in trials • Discussion at stakeholder meetings • Inclusion in quality improvement initiatives / standards of care • Inclusion in guidelines on clinical evaluation of interventions
Patient/ consumers	• Include a link to the SONG Website • Patient newsletter

	• Advocacy
Industry	• Include in protocols
Researchers	• Targeted correspondence and meetings with triallists • Use core outcomes in trials/research

5.2 Endorsement from stakeholder organisations

For each set of core outcome domains and/or measure, the Executive Committee/Steering Group/Expert Working Group will invite stakeholder organisations to provide endorsement. A summary article (e.g. in the form of a commentary) that details the process and results will be drafted and sent to stakeholder organisations. Any endorsements will be listed with the commentary and submitted to a relevant journal or journals.

5.3 Publication in peer-reviewed journals

The protocol and all phases of the SONG streams are submitted to publication in peer-reviewed biomedical journals. The choice of journals is decided by the relevant Committee/Group.

5.4 SONG newsletter

The SONG e-newsletter provides an update on SONG-related activities and highlights opportunities for stakeholders to get involved. The newsletter is sent by email to the database. Copies of e-newsletter can be accessed via http://songinitiative.org/news/

5.4 Evaluation

Short-term: Mixed methods processes evaluation studies may be conducted to evaluate the implementation of the core outcomes in trials.

Long-term: OMERACT conducted an observational review of 350 randomised trials for the treatment of rheumatoid arthritis to ascertain the use of the core outcome set. This review demonstrated improvements in the consistency of measurement of the core outcome set since it was introduced two decades ago[26]. A similar review may be conducted after the core outcome set is identified.

6 | SONG Nomenclature

This section lists the terms used throughout the SONG process.

Table 6a. Category of outcomes

Category	Definition	Examples
Surrogate/biochemical	A surrogate endpoint or outcome is a biochemical, imaging, or other marker used as a substitute for a clinical outcome[27]. Ideally, the surrogate should exist within the therapeutic pathway between the intervention and meaningful benefit.	Potassium, calcium, phosphate
Clinical	A medical event or comorbidity diagnosed by the clinician[28].	Mortality, cardiovascular disease, hospitalisation
Patient-reported	An outcome reported directly from patients regarding how they function or feel in relation to a health condition and its therapy, without interpretation by a healthcare professional or anyone else[29].	Fatigue, pain

Table 6b. Classification of outcomes and definitions

Level	Definition	Examples
Outcome domain	A broad term to include a set of specific outcomes/outcome measures	Mortality, cardiovascular disease, fatigue, vascular access
Outcome/dimension*	The specific health outcome/impact of an intervention.	Myocardial infarction, vascular access function, impact of fatigue on life participation*
Outcome measure	The specific measure/instrument, definition/threshold: specific measurement (e.g. the name of the scale used), specific metric (to characterise the patient's results such as change in baseline at time X), method of aggregating the data (e.g. mean or median for continuous or proportion for categorical measures).	Proportion of patients with fatal myocardial infarction, change from baseline. FACIT-F (survey to measure fatigue), end value, proportion of patients with a score of > 45.
Time point	The time at which the outcome is measured.	12 months
Original outcome**	The descriptor used in the original trial which may include the metric and method of aggregating the data (this is not always reported at the same level of detail)	-

*Dimension may be more appropriate for patient-reported outcomes; **For systematic reviews only

Glossary

Adverse event An unintended consequence and sometimes harmful occurrence in a patient that may or may not be associated with the intervention given to them. This may not be a side effect because it is not always clear if the intervention caused the effect.

Best worst scale A survey that assesses the relative importance of items (i.e. outcomes).

Caregiver Family members or friends involved in the care of the patient. They provide care in an unpaid and voluntary capacity.

Chronic kidney disease An abnormality in the kidneys that is present for more than three months.

Clinical outcome A medical event or comorbidity diagnosed by the clinician.

Clinical trial A type of study that is used to evaluate and compare the effect of two or more interventions and is usually conducted in patients. *See also Randomised controlled clinical trial.*

COMET The Core Outcome Measures in Effectiveness Trials Initiative that facilitates the development and application of core outcome sets www.comet-initiative.org/

Consensus Refers to agreement. In the context of SONG, this is agreement among stakeholder groups – patients/caregivers and health professionals.

Core outcome domains A broad term to include a set of specific outcomes/outcome measures that should be reported in all trials in the specific treatment or CKD stage.

Core outcome measure The specific measure/instrument, definition/threshold, specific metric (to characterise the patient's results such as change in baseline at time X), method of aggregating the data (e.g. mean or median for continuous or proportion for categorical measures).

Core outcome set An agreed minimum set of outcomes or outcome measures. It is a recommendation of 'what' should be measured and reported in all trials in a specific area.

Delphi survey A technique for achieving consensus among a panel of experts. This usually involves two to three rounds of surveys, answered anonymously by participants. Participants can reflect on the scores of previous rounds to inform their responses in subsequent rounds.

Dialysis A treatment for end stage kidney disease that removes waste products and excess fluid from the blood by filtering the blood through a membrane. *See also Haemodialysis and Peritoneal dialysis.*

End-stage kidney disease A term for advanced kidney failure. Patients with end-stage kidney disease usually need dialysis or kidney transplantation to survive.

Haemodialysis A type of treatment for kidney failure that uses a machine to filter the patient's blood.

Intervention Something that is done (e.g. medications, program, surgical procedure, strategy, policies) in an effort to improve patient health. In trials, the intervention is a treatment of other health care course of action under investigation.

Kidney transplant A procedure that involves transplanting a kidney from a donor into a patient who needs a kidney transplant. The transplanted kidney may also be called a graft.

Nephrologist A kidney specialist.

Nephrology A specialised area of medicine that is focussed on the kidneys.

Nominal group technique A small group discussion designed to achieve consensus. Participants submit ideas (e.g. outcomes), discuss their choices, and prioritise the ideas (outcomes).

OMERACT The Outcomes Measures in Rheumatology is an independent initiative established to develoip core outcomes and outcome measures for patients with musculoskeletal conditions.

Outcome Researchers look at the effects those treatments have on patients and do this by measuring an "outcome". An outcome is something that can be measured, and can arise or change because of a health condition or treatment. This may also be called an endpoint. *See also Primary outcome and Secondary outcome.*

Patient A person diagnosed with a disease (e.g. kidney disease) or has to make a health-related decision for themself.

Patient-reported outcome An outcome reported directly from patients regarding how they function or feel in relation to a health condition and its therapy, without interpretation by a healthcare professional or anyone else. Patient-reported outcome measures (PROMS) (e.g. surveys) are completed by patients to assess these patient-reported outcomes.

Peritoneal dialysis A type of treatment for kidney failure that uses the lining of the abdomen (peritoneum) and a solution (dialysate) to clean the blood.

Primary outcome The outcome that a researcher (or investigator) considers to be the most important outcome for the specific trial. This needs to be identified and defined at the time of designing the study.

Protocol Detailed plan for a study.

Randomised controlled clinical trial A trial in which two or more interventions (usually includes a control intervention, no intervention, placebo) are compared. The researchers randomly allocate participants to groups (e.g. treatment and control) and compare the effect of outcomes between the groups.

Sample size The number of participants in a trial.

Secondary outcome An outcome that may be used to evaluate other effects of the intervention.

Side effect An unintended, unexpected or undesirable result of an intervention.

Stakeholder (including stakeholder organisation) In the context of SONG, an individual (or organisation) with an interest in core outcomes for trials in chronic kidney disease.

Surrogate outcome A surrogate endpoint or outcome is a biochemical, imaging, or other marker used as a substitute for a clinical outcome. Ideally, the surrogate should exist within the therapeutic pathway between the intervention and meaningful benefit.

Systematic review A comprehensive and structured review that involves collecting and analysing multiple research studies or articles.

References

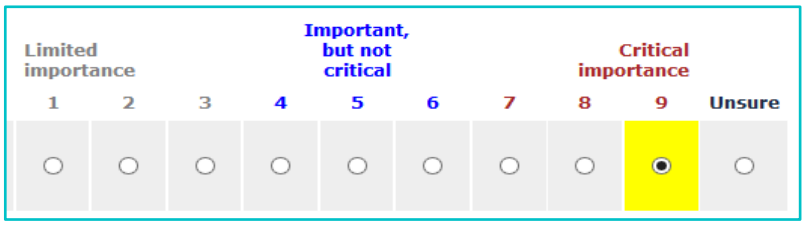
1. Yudkin JS, Lipska KJ, Montori VM. The idolatry of the surrogate. *Br Med J*. 2011;343:d7995.
2. Williamson PR, Altman D, Blazeby JM, et al. Developing core outcomes sets for clinical trials: issues to consider. *Trials*. 2012;13:132.
3. Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ. Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ*. 2014;349:g6501.
4. Kirkham JJ GS, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, Moher D, Schmitt J, Tugwell P, Tunis S, Williamson PR. Core Outcome Set-STAndards for Reporting: The COS-STAR Statement. *PLoS Med*. 2016;13(10):e1002148.
5. Prinsen CA, Vohra S, Rose MR, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials*. 2016;17(1):449.
6. Howell M TA, Wong G, Craig JC, Howard K. Important outcomes for kidney transplant recipients: a nominal group and qualitative study. *Am J Kidney Dis*. 2012;60(2):186-196.
7. Urquhart-Secord R CJ, Hemmelgarn B, Tam-Tham H, Manns B, Howell M, Polkinghorne KR, Kerr PG, Harris DC, Thompson S, Schick-Makaroff K, Wheeler DC, van Biesen W, Winkelmayer WC, Johnson DW, Howard K, Evangelidis N, Tong A. Patient and caregiver priorities for outcomes in hemodialysis: an international nominal group technique study. *Am J Kidney Dis*. 2016;68(3):444-454.
8. Corner J WD, Hopkinson J, Gunaratnam Y, McDonald JW, Foster C. The research priorities of patients attending UK cancer treatment centres: findings from a modified nominal group study. *Br J Cancer*. 2007;96(6):875-881.
9. Jones J, Hunter D. Consensus methods for medical and health services research. *Br Med J*. 1995;311(7001):376-380.
10. Dalkey N, Helmer O. An experimental application of the Delphi method to the use of experts. *Management Science*. 1963;9(3):458-467.
11. Chiarotto A TC, 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.
12. Schunemann H BJ, Oxman A. GRADE Handbook for grading quality of evidence and strength of recommendation available at http://gdt.guidelinedevelopment.org/app/handbook/handbook.html. 2009.
13. Flynn TN LJ, Peters TJ, Coast J. . Best–worst scaling: What it can do for health care research and how to do it. *J Health Econ*. 2007;26(1):171-189.
14. Louviere JJ, Flynn TN LJ, Peters TJ, Coast J. Using best-worst scaling choice experiments to measure public perceptions and preferneces for healthcare reform in Australia. *Patient*. 2010;3(4):275-283.
15. Tong A MB, Hemmelgarn B, Wheeler DC, Evangelidis N, Tugwell P, Crowe S, Van Biesen W, Winkelmayer WC, O'Donoghue D, Tam-Tham H, Shen JI, Pinter J, Larkins N, Youssouf S, Mandayam S, Ju A, Craig JC. Establishing core outcome domains in hemodialysis: report of the standardized outcomes in nephrology - hemodialysis (SONG-HD) consensus workshop. *Am J Kidney Dis*. 2017;69(1):97-107.
16. Boers M, Kirwan RJ, Tugwell P, et al. *The OMERACT Handbook*. 2017.
17. Rothrock NE, Kaiser KA, Cella D. Developing a valid patient-reported outcome measure. *Clin Pharmacol Ther*. 2011;90(5):737-742.
18. Flnk A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. *Am J Public Health*. 1984;74(9):979-983.
19. Drennan J. Cognitive interviewing: verbal data in the design and pretesting of questionnaires. *J Adv Nurs*. 2003;42(1):57-63.
20. Morgan SJ, Amtmann D, Abrahamson DC, Kajlich AJ, Hafner BJ. Use of cognitive interviews in the development of the PLUS-M item bank. *Qual Life Res*. 2014;23(6):1767-1775.
21. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technol Assess*. 1998;2(14):1-86.
22. King MT. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoeconom Outcomes Res*. 2011;11(2):171-184.
23. FDA. US Food and Drug Administration. Volume I - 5.4 Test Methods and Method Validation available at: https://www.fda.gov/ScienceResearch/FieldScience/ucm171818.htm (Accessed 25th April 2017).

24. DDL. What is test method validation (TMV) and what can it do for me? Available at https://www.testedandproven.com/wp-content/uploads/2015/04/DDL.TMV_.Whitepaper.4.23.15.pdf (Accessed 25th April 2017).
25. Evangelidis N, Tong A, Manns B, et al. Developing a set of core outcomes for trials in hemodialysis: an international Delphi survey. *Am J Kidney Dis*. 2017:doi: 10.1053/j.ajkd.2016.1011.1029 (online first).
26. Kirkham JJ, Boers M, Tugwell P, Clarke M, Williamson PR. Outcome measures in rheumatoid arthritis randomised trials over the last 50 years. *Trials*. 2013;14:324.
27. Aronson JK. Biomarkers and surrogate endpoints. *Br J Clin Pharmacol*. 2005;59(5):491-494.
28. Velentgas P, Dreyer NA, Wu AW. Outcome Definition and Measurement. In: Velentgas P DN NP, ed. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. United States: Agency for Healthcare Research and Quality (2013.
29. Higgins JPT, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Collaboration;2011.
30. Zarin DA TT, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database--update and key issues. *N Engl J Med*. 2011;364(9):852-860.
31. Velentgas P, Dreyer NA, Wu AW. *Chapter 6. Outcome definition and measurement*. Washington DC, United States: Agency for Healthcare Research and Quality;2013.

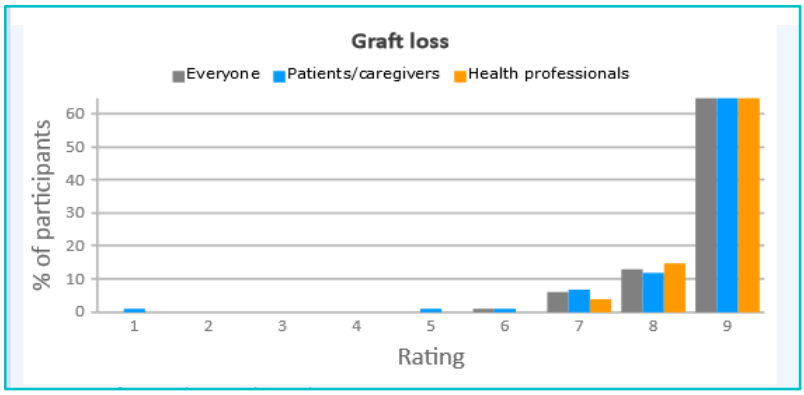
Appendices

Appendix 2a. Screenshot of a SONG Delphi Survey

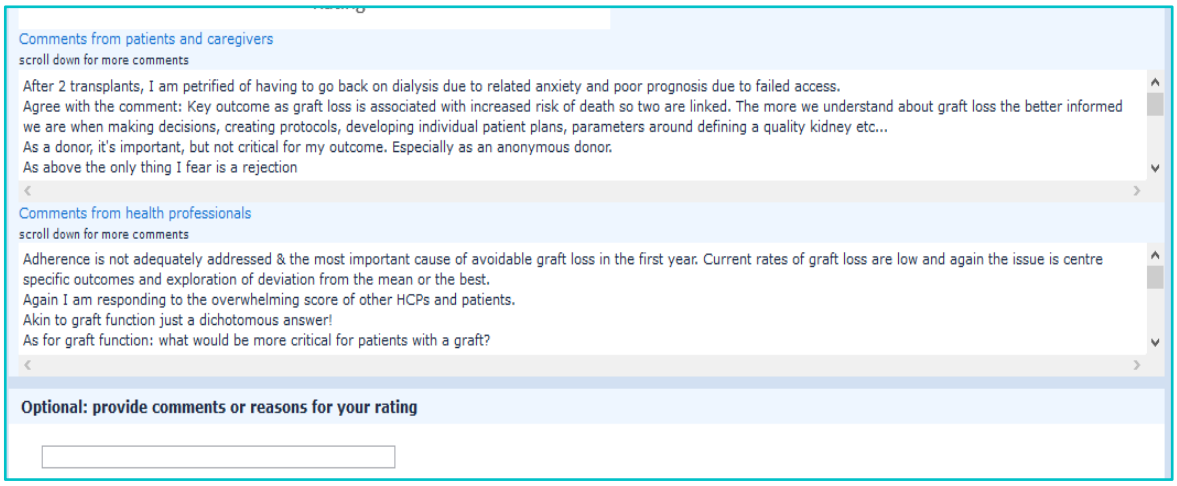
Rating scale



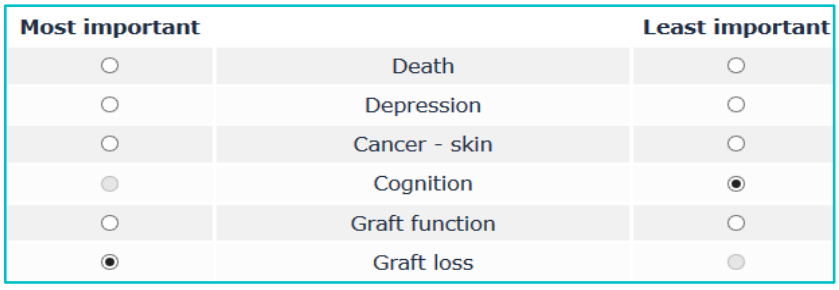
Distribution graph



Free text comments



Best-Worst Scale



Appendix 2b. Explanatory notes on the inclusion of mortality as a core outcome domain

A core outcome set for clinical trials should capture both quality and quantity of life to comprehensively assess impact of interventions on health. This is supported by the Outcome Measures in Rheumatology (OMERACT) initiative, which has improved the reporting and relevance of outcomes in rheumatology trials. OMERACT recommends the inclusion of mortality alongside 'life impact' (e.g. quality of life dimensions such as fatigue). Core outcomes sets in other health areas (e.g. pregnancy and birth, hip fracture, asthma, low back pain) also include mortality, health conditions and complications, and quality of life.

Mortality is infrequently reported and inconsistently measures in trials.

Appendix 3a. Framework for outcome measures and examples

Zarin 2011 NEJM[30]

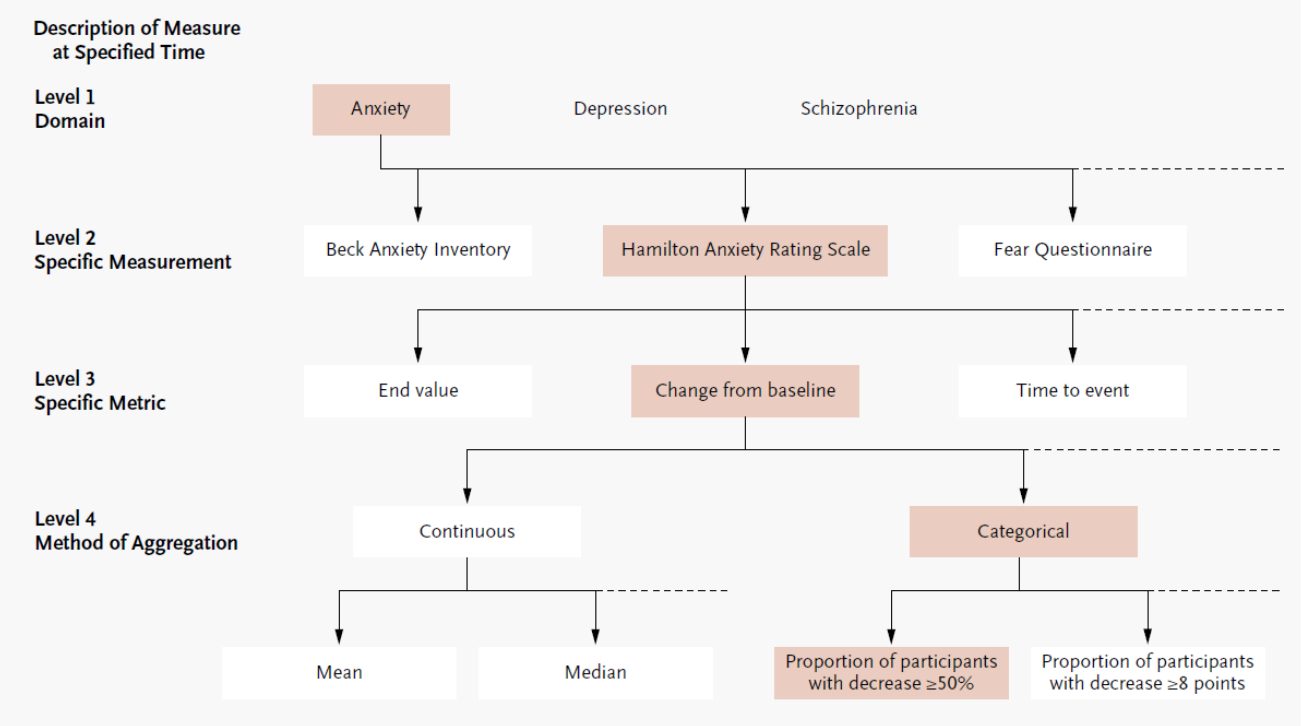


Figure 1. An Example of the Four Levels of Specification in Reporting Outcome Measures.

AHRQ 2013[31]

Table 6.2. Clinical outcome definitions and objective measures

Conceptual	Temporal Aspects	Objective Measure
Incident invasive breast cancer	Incident	SEER or state cancer registry data
Myocardial infarction	Acute, transient (in regard to elevated Troponin-I)	Review of laboratory test results for troponin and other cardiac enzymes for correspondence with a standard clinical definition
Psoriasis	Chronic, prevalent	Psoriasis Area Severity Index (PASI score) or percent body surface area assessment
Systematic lupus erythematosus (SLE)	Chronic condition with recurrent flares (Episodes may have acute onset)	Systemic Lupus Erythematosus Disease Activity Index (SLEDAI)

Clinicaltrials.gov Outcome Measure Template

Outcome Measure Template	*Example 1*	(Units=Participants; Measure Type=Number; Measure of Dispersion/Precision=Not Applicable)	*ClinicalTrials.gov*

* Outcome Measure Type	*(Circle One)* (Primary)	Secondary	Other Pre-specified	Post-Hoc	**Safety Issue?**	*(Circle One)* (Yes) No
* Outcome Measure Title	**Number of Participants With Myocardial Infarction, Stroke or Death From Cardiovascular Causes**					
Outcome Measure Description	Participants were monitored for up to 2 years. This is the number of participants who have had at least one myocardial infarction or stroke, or if they died from cardiovascular causes during the time of observation.					
* Outcome Measure Time Frame	Up to 2 years					

* Outcome Measure Type	*(Circle One)* (Primary)	Secondary	Other Pre-specified	Post-Hoc	**Safety Issue?**	*(Circle One)* Yes (No)
* Outcome Measure Title	**Percentage of Participants Achieving Predefined Antibody Level ≥0.1 International Units Per Milliliter (IU/mL) for Tetanus Toxoid**					
Outcome Measure Description	Percentage of participants achieving predefined antibody threshold ≥0.1 IU/ mL along with the corresponding 95% CI for concomitant antigen tetanus toxoid are presented. Exact 2-sided CI was based on the observed proportion of participants.					
* Outcome Measure Time Frame	1 month after the infant series (7 months of age)					

* Outcome Measure Type	*(Circle One)* (Primary)	Secondary	Other Pre-specified	Post-Hoc	**Safety Issue?**	*(Circle One)* Yes (No)
* Outcome Measure Title	**Change in Low-density Lipoprotein (LDL) Cholesterol**					
Outcome Measure Description						
* Outcome Measure Time Frame	Baseline, 3 months					

* Outcome Measure Type	*(Circle One)* (Primary)	Secondary	Other Pre-specified	Post-Hoc	**Safety Issue?**	*(Circle One)* Yes (No)
* Outcome Measure Title	**Median Time to Response of Target Lesions using RECIST Criteria**					
Outcome Measure Description	Median Time from 1st dose of treatment to Complete or Partial Response. Target lesions are scanned via MRI to determine dimensions. Response Evaluation Criteria In Solid Tumors (RECIST) Complete Response is defined to be a disappearance of all target lesion(s). RECIST Partial Response is defined to be at least a 30% decrease in the sum of the target lesion longest diameters (LDs).					
* Outcome Measure Time Frame	Up to 24 months					

Appendix 3b. Examples of outcome measures for SONG-HD Vascular Access

Here are some outcome measures (including examples with metrics/method of aggregation) that have been used or suggested by the Expert Working Group to assess specific vascular access outcomes. "Function" and "Infection" were chosen as examples as they were the top prioritised vascular access outcomes based on the preliminary results of the survey.

Outcome measures to assess "Function"
• Ability to use the vascular access for haemodialysis *(e.g. the proportion of times the vascular access can be used for haemodialysis)*
• Uninterrupted use of the vascular access for haemodialysis without need for any access intervention/procedures *(e.g. the average number of days the vascular access can be used for haemodialysis without interruptions for interventions [e.g. de-clotting, line change])*
• Relative change in access blood flow *(e.g. percentage change in access blood flow)*
• Ability to receive two needle cannulation to achieve prescribed dialysis within a given timeframe *(e.g. the proportion of times the patient is able to receive the prescribed dialysis using 2 needle cannulation within a month)*

Outcome measures to assess "Infection"
• Bloodstream infection in the absence of any other identifiable cause than the vascular access *(e.g. the number of episodes of infection from the vascular access detected in the bloodstream per 1000 access days)*
• Definite or presumed local (around the access site) or systemic (in the bloodstream) vascular access infection requiring antibiotic treatment *(e.g. number of episodes of vascular access infections around the access site or in the bloodstream requiring antibiotic treatment per 1000 access days)*

Examples of measurement properties – function

Property	Simplified definition	Example of an outcome measure to assess the function of a vascular access: "relative change in access blood flow"
Content validity	Whether the outcome measure captures all the relevant aspects of the outcome.	Does measuring a change in blood flow capture all the relevant aspects of vascular access function, i.e. how well the vascular access is working?
Reliability	Whether measuring the outcome provides similar or the same result on multiple occasions under consistent conditions.	If the blood flow of a vascular access is measured several times during a haemodialysis session, are all the flow results similar or the same?
Responsiveness	How well a change in the outcome over time can be detected by the outcome measure being used.	If the function of a vascular access gets worse over time, does a relative change in blood flow measured at different time points demonstrate this change?

Internal consistency	How closely different measures that assess the same outcome are related.	If function is measured by "relative change in access blood flow" and by "absence of thrombosis", how much are blood flow and thrombosis (blood clot) related?
Measurement error	Changes in measurement results that are not due to true changes in the outcome but related to problems with accurate measurement.	If the dialysis needles are not placed correctly, the measured blood flow will be inaccurate/wrong.
Criterion validity	Whether the outcome measure is a good reflection of the "gold standard" or best instrument available for that measurement (provided that there is one).	Is the access blood flow measured by ultrasound dilution technique an accurate/ good reflection of the true blood flow if measured with a flow probe inside the vessel?
Cross cultural validity	Whether an outcome measure can be used across cultures and reproduce the same results.	Does the technique and equipment used to measure vascular access blood flow vary across cultures and countries?

Examples of feasibility aspects – function

Property	Examples of an outcome measure to assess the function of a vascular access:: "relative change in access blood flow"
The patient can understand it	"Relative change in access blood flow" versus "a change in the amount of blood that flows through the vascular access"
We can interpret the results	"a 25% decrease in access blood flow"
It is easy to administer	Measuring access blood flow needs equipment such as a special ultrasound machine and experienced staff to perform the measurement correctly.
Is it easy to standardise?	When should the access blood flow be measured, how often and what relative change is considered clinically relevant?
The time it takes to complete	Measuring access blood flow takes between approximately 5 min and 30 min depending on the equipment and method used.
The clinician can understand it	"Relative change in access blood flow" versus "Percentage change in blood flow over a certain period of time"
The cost	The costs involved for staff to measure the blood flow and the equipment (~24-27,000 USD) required to measure the blood flow.
The equipment needed	A special ultrasound machine
It is available in different settings	The special ultrasound machine to measure the blood flow may not be available in every dialysis unit.
How invasive it is	Blood flow assessments are usually performed during a haemodialysis session which is an invasive procedure.
How easy it is to assess the results	It involves comparison of blood flow measures to previous blood flow measures.