Sinclair, S., Jaggi, P., Hack, T. F., Russell, L., McClement, S. E., Cuthbertson, L., Selman, L. E., & Leget, C. (2020). Initial validation of a patient-reported measure of compassion: Determining the content validity and clinical sensibility among patients living with a life-limiting and incurable illness. *Patient*, [11]. https://doi.org/10.1007/s40271-020-00409-8

Initial validation of a patient-reported measure of compassion: Determining the content validity and clinical sensibility among patients living with a life-limiting and incurable illness.

Shane Sinclair PhD,[1,2] Priya Jaggi MSc,[1] Thomas F. Hack PhD,[3,4,5] Lara Russell PhD,[6,7] Susan E. McClement PhD,[3] Lena Cuthbertson,[8] Lucy E Selman PhD,[9] Carlo Leget PhD[10]

1 Faculty of Nursing, University of Calgary, 2500 University Drive NW Calgary, Alberta Canada. T2N 1N4.
2 Department of Oncology, Cumming School of Medicine, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada. T2N 1N4.
3 College of Nursing, Rady Faculty of Health Sciences, University of Manitoba, 89 Curry Place, Winnipeg, Manitoba, Canada. R3T 2N2.
4 Research Institute in Oncology and Hematology, CancerCare Manitoba, 4005E – 675 McDermot, Winnipeg, Manitoba, Canada. R3E 0V9.
5 Psychosocial Oncology and Cancer Nursing Research, IH Asper Clinical Research Institute, 369 Taché Ave, Winnipeg MB Canada R2H 2A6.
6 Centre for Health Evaluation and Outcome Sciences, 588 – 1081Burrard Street, Vancouver, B.C., V6Z 1Y6.
7 Trinity Western University, 7600 Glover Road, Langley, BC, Canada V2Y 1Y1
8 Office of Patient-Centred Measurement, British Columbia Ministry of Health, c/o Providence Health Care, 1190 Hornby Street, 3rd Floor, 341F, Vancouver, British Columbia, Canada. V6Z 2K5.
9 Population Health Sciences, Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, United Kingdom. BS8 2PS.
10 University of Humanistic Studies, Kromme Nieuwegracht 29, 3512 HD Utrecht, The Netherlands.

Corresponding Author: Dr. Shane Sinclair
Tel: 403.220.2925
E-mail: sinclair@ucalgary.ca
Fax: 403.284.4803

Co-Authors:
Priya Jaggi (parneetpriya.jaggi@ucalgary.ca)
Thomas F. Hack (thack@sbrc.ca)
Lara Russell (lrussell@cheos.ubc.ca)
Susan E. McClement (Susan.Mcclement@umanitoba.ca)
Lena Cuthbertson (lcuthbertson@providencehealth.bc.ca)
Lucy Selman (lucy.selman@bristol.ac.uk)
Carlo Leget (C.Leget@UvH.nl)

Abstract

Background: Although compassionate care is considered a cornerstone of quality palliative care, there is a paucity of valid and reliable measures to study, assess, and evaluate how patients experience compassion/compassionate care in their care.

Objective: To develop a patient-reported compassion measure for use in research and clinical practice with established content-related validity evidence for the items, question stems, and response scale.

Methods: Content validation for an initial 109 items was conducted through a two-round modified Delphi technique, followed by cognitive interviews with patients. A panel of international, Subject Matter Experts (SMEs) and a Patient Advisory Group (PAG) assessed the items for their relevancy to their associated domain of compassion, yielding an Item-level Content Validity Index (I-CVI), which was used to determine content modifications. The SMEs and the PAG also provided narrative feedback on the clarity, flow, and wording of the instructions, questions, and response scale, with items being modified accordingly. Cognitive interviews were conducted with 16 patients to further assess the clarity, comprehensibility, and readability of each item within the revised item pool.

Results: The first round of the Delphi review produced an overall CVI of 72% among SMEs and 80% among the PAG for the 109 items. Delphi panelists then reviewed a revised measure containing 84 items, generating an overall CVI of 84% for SMEs and 86% for the PAG. Sixty-eight items underwent further testing via cognitive interviews with patients, resulting in an additional 14 items being removed.

Conclusions: Having established this initial validity evidence, further testing to assess internal consistency, test-retest reliability, factor structure, and relationships to other variables is required to produce the first valid, reliable, and clinically informed patient-reported measure of compassion.

Key Words: Content Validity, Compassion, Clinical sensibility, Psychometric properties, Patient-reported measure

Key Points for Decision Makers:

- Compassion, is recognized as essential factor of quality care, that needs to be measured in order to improve care and conduct high-quality research.
- Compassion measures in healthcare exist, however their reliability and validity is limited, including their content-related validity evidence, significantly diminishing their overall psychometric soundness and clinical utility,
- A patient-orientated measure of compassion with content-related validity evidence will provide researchers, clinicians, and decision makers with the means to assess, analyze, and improve factors associated with compassion within healthcare.

1 Introduction

Patient-Reported Experience Measures (PREMs) are increasingly recognized as a valid and reliable means of assessing patients' overall experience of clinical care [1-3]. While Patient-Reported Outcome Measures (PROMs) give patients the opportunity to provide self-reports of their well-being or health-related quality of life in general and the outcomes of care or impact of interventions, specifically, PREMs capture patients' overall satisfaction and experiences of the processes of care, as they are receiving it, indicating areas of exemplary care and those needing improvement [1-3]. Compassion, "a virtuous response that seeks to address the suffering and needs of a person through relational understanding and action" [4], p.195, is identified by patients as a pillar of palliative care, yet evidence suggests this essential ingredient is poorly addressed by healthcare providers and healthcare systems in general [5-9].

Efforts to improve compassion in healthcare are hindered by a lack of evidence-based measures of compassion [4, 5, 9-12]. A recent systematic review of existing compassion measures concluded there is "no single measure that measured the construct in a comprehensive or sufficiently methodologically rigorous fashion" [13], p.404. Of the nine tools identified to measure compassion [13], each has significant limitations, including but not limited to the lack of patient perspectives and lack of evidence of construct validity, as revealed by EMPRO (Evaluate the Measurement of Patient-Reported Outcomes) scores for each measure ranging from 4.76 – 28.57 (out of 100) [13]. For example, although the Compassion Competence Scale assesses a number of core elements of compassion, it is a self-report of healthcare providers' compassion, as opposed to patient ratings [13]. The Schwartz Centre Compassionate Care Scale measures patients' perceptions of compassionate care provided by hospital physicians, and although it received the highest EMPRO score, it only assesses limited domains of compassion [13]. These, along with various other methodological limitations, have left clinicians, educators, researchers, and health system leaders without the means to empirically evaluate and improve a central aspect of quality care [13].

To address these limitations, we first conducted a rigorous, grounded theory, qualitative investigation of advanced cancer patients' understandings and experiences of compassion, generating an empirical Patient Compassion (PCM). The PCM delineates the key components of compassion and their relationship to one another, including delineating it from sympathy and empathy [4, 14]. With a foundational model of compassion solely from the patient perspective, we then embarked on a multi-centre study to develop and validate a patient-reported compassion measure with a two-fold purpose: a) to measure patients' *experience of compassion* based on the

emanated behaviours, skills, and qualities of their Healthcare Providers (HCPs); and b) to determine the extent to which patients feel that the care they received was compassionate [15]. The purpose of developing a compassion measure was to provide clinicians the means to measure individual patient experiences of compassion, while also providing researchers with a gold standard measure to conduct high-quality compassion research in healthcare. Following measure development guidelines [16-18], the compassion measure was developed as a self-reported experience measure for individuals living with an incurable, life-limiting illness across diverse patient populations and care settings, including acute palliative care, hospice, palliative home care and long-term/residential care. The objective of this study is to examine the content-related validity evidence for the initial items of this newly developed compassion measure utilizing Subject Matter Experts (SMEs), members of a Patient Advisory Group (PAG), and cognitive interviews with patients within the aforementioned care settings.

Figure 1. Patient Compassion Model [4].

2 Methods

Prior to establishing the content validity evidence described herein, we assessed the face validity, credibility and transferability of the PCM (Figure 1) [19]. Next, seven members of the research team systematically generated initial candidate items utilizing the themes, categories, and codes of the PCM (Figure 1), Following five phases of measure development [15], the candidate items were circulated amongst the research team in additional iterations, with each member providing in-depth, written feedback on the items and their potential response scales [15]. The result was a revised item pool of 109 items [15] covering each of the domains of compassion within the PCM. This manuscript reports the content-related validity evidence of the compassion measure by assessing content relevancy, content representativeness, and content quality [20-23]. The goal was to establish a pool of relevant, representative, and clear items with an appropriate response scale. Clinical sensibility testing, which focuses on how well a measure addresses the topic of interest [24], was also assessed amongst patients to ensure the clarity, readability, and wording of the instructions, questions and response scales of the proposed measure.

2.1 Study Population and Data Collection:

2.1.1 Modified Delphi: Subject Matter Expert (SME) and Patient Advisory Group (PAG) participants

Consensus on a draft version of the compassion measure (109 items spanning 6 domains) was established through two rounds of a modified Delphi technique [25, 26] with Subject Matter Experts (SMEs)

and a Patient Advisory Group (PAG). A list of potential SMEs was generated from our previous literature reviews [13, 27] and snowball sampling, whereby potential SMEs were asked to recommend colleagues. The inclusion criteria for SMEs was: English-speaking, at least an MD or PhD qualification, and at least 5 years of academic research on the topic of compassion. Thirteen international SMEs were identified and invited via email to participate in the study via an online survey.

Recognizing the aforementioned limitations of previous measures and that the aim of the study was to develop a patient-reported experience measure, we felt it was imperative to include the patient perspective across all study stages through a Patient Advisory Group (PAG). Twenty patients were recruited from the Alberta Cancer Foundation, Patient Partnerships, and the Alberta Innovates, Strategy for Patient-Oriented Research SUPPORT Unit [28].

Survey data was gathered via REDCap [29] and was designed to assess the following aspects of the compassion measure: 1) item relevancy; 2) representativeness of the items collectively to their respective domains; 3) item clarity; 4) the relevancy of the recall period and the response scale; and 5) the perceived importance of a patient-reported measure of compassion in informing patient care. Item relevancy, *that items reflect, sample, and measure the domain of interest* [30, 31], was assessed by having participants review each of the 109 items in the draft measure (*item-level*) for their relevancy to their respective domain of compassion [4] on a 4-point Likert scale ranging from 1=not relevant to 4=highly relevant. Representativeness, *the degree to which the collective items together cover the larger domain of interest* [31], was also assessed on a 4-point Likert Scale ranging from 1=not representative to 4=highly representative, by participants, along with the opportunity to provide narrative feedback on items that were particularly problematic or items they felt should be added to improve domain representativeness. Item clarity was assessed, by asking participants whether each item was clear (Yes/No), with the opportunity to provide open feedback on suggested re-wording. All participants were also asked to indicate on a 5-point Likert scale their level of agreement (ranging from 1=strongly disagree to 5=strongly agree), with the proposed recall period; the relevancy of the response scale; and to appraise the relevancy of the collective items to their respective domain and the overarching construct of compassion. The PAG was also asked an additional question to assess the perceived importance of the measure to patient care, ranging from 1= not at all important and 5=extremely important.

Finally, because we intentionally included a number of alternatively worded items, participants were asked to indicate their preference between alternate items.

2.1.2 Clinical Sensibility Testing: Cognitive interviews with patient participants

A cognitive interview guide (Appendix A), was administered by a trained Research Assistant to further assess the measure items that were retained after the two Delphi rounds [32]. Utilizing purposive sampling, we recruited a diverse sample of patients living with a life limiting illness from acute care, home care, hospice, and long term care within a large urban setting in Western Canada. Participants were asked to use a 'think-aloud' process, verbalizing their free thoughts and personal interpretation of each item [32]. Comprehension, recall, and judgment of the items, along with response categories and question stems, were assessed via probes [32] (Appendix A). Participants were also asked to indicate their preference between a number of alternatively-worded items from the Delphi stage that required further assessment.

Data was collected between August, 2017 and January, 2018. This study was approved by the University of Calgary Conjoint Health Research Ethics Board (REB #17-0754). Written informed consent to participate was obtained from all participants.

2.2 Data Analysis*:*

2.2.1 Modified Delphi: Item, Domain, and Construct-Level Content Validity Index

Data analyses were conducted using IBM SPSS Statistics Software, Version 24. A Content Validity Index, a measure of consensus that indicates the degree of response convergence between respondents regarding the content relevancy of a measure, for the individual items (I-CVI) was calculated by dividing the number of items given a rating of either 3 (quite relevant) or 4 (highly relevant) by the total number of experts [33]. Through a process of a consensus, the research team pre-determined I-CVI cut-off levels (Appendix B), based on the number of experts that participated per round. These I-CVI cut-off levels facilitated our decision-making on whether individual items should be candidates for reconsideration (i.e. modification) or deletion, and also helped to flag items that were viewed as "problematic (Appendix B).

SME and PAG I-CVIs and raw qualitative data obtained from each Delphi round were summarized in tabular form and assessed independently by 2 reviewers (PJ and MH) who made an initial recommendation to either keep, discard, modify or flag each item for further discussion. Disagreements were resolved through discussion with the Principal Investigator (SS). In addition to the I-CVIs, and the

independent reviewers' recommendations, the larger research team was provided the Scale Content Validity Index (S-CVI/Ave), calculated by dividing the sum of the I-CVIs by the total number of items in the entire scale or within individual domains, in order to identify domains that required improvement [33]. The final pool of items to be assessed through cognitive interviews was determined through a series of videoconferences with members of the research team, on an item-by-item basis, until consensus was reached. An *a priori* criterion S-CVI of 80% of the items collectively deemed as relevant in the entire compassion measure was utilized as the cut-off for determining whether the measure's items were relevant as a whole, or whether additional rounds of review would be required [34, 35].

Median ratings for domain representativeness, and relevancy of the proposed recall period and response scale were also assessed in each round of review. Higher median ratings in conjunction with a decrease in the rating range (i.e. lower variance) were used to indicate improvement with these aspects of interest over the rounds of review.

2.2.2. Cognitive interviews: Qualitative analysis

Qualitative data from the cognitive interviews were recorded, analyzed, collated and summarized in tabular form in accordance with the framework analysis method [36]. The matrix-based approach facilitated the development of a framework, with each row consisting of the items and the columns for each participant. Two members of the research team (SS and PJ) independently reviewed each interview and recorded issues for each question related to comprehension, recall, judgment, response categories, and question stems. The reviewers compared and contrasted their results, coming to consensus on differences through an iterative process of discussion prior to collating their independent results into a master framework, which was then circulated to the research team for final consideration. Three analysis meetings with the larger research team occurred after members had independently reviewed the results. Each item and its associated participant response was reviewed by the research team, with members indicating their agreement with the proposed suggestions until consensus was achieved.

3 Results

3.1 Participant Characteristics: The Delphi Panel

Of the 8 SMEs who agreed to participate in the 1st Delphi round, 6 surveys were completed, with 2 incomplete surveys excluded. In round two, a total of 8 SMEs participated, as 2 additional SMEs were

identified via snowball sampling. From the 20 patient advisors who expressed interest in participating in this study, 9 were selected to form the PAG (Table 1).

Table 1. Demographics: Subject Matter Expert (n=8) and Patient Advisory Group (n=9).

Characteristic	Subject Matter Experts (n=8)	Patient Advisory Group (n=9)
	Frequency (Percentage)	Frequency (Percentage)
Profession		
Physician	2 (25%)	---
Registered Nurse	4 (50%)	---
Spiritual Care Specialist	1 (12.5%)	---
Psychologist	1 (12.5%)	---
Highest, Completed Level of Education		
PhD	6 (75%)	---
MD	2 (25%)	---
Some University or Technical school	---	2 (22.2%)
University/College/Technical school completed	---	4 (44.4%)
Post graduate University completed	---	3 (33.3%)
Gender		
Male	5 (62.5%)	4 (44.4%)
Female	3 (37.5%)	5 (55.5%)
Religious Group Affiliation		
Christian	5 (62.5%)	5 (55.5%)
Buddhist	1 (12.5%)	---
Hindu	1 (12.5%)	---
Other	---	1 (11.1%)
None	1 (12.5%)	3 (33.3%)
Religiousness or spirituality		
Spiritual and religious	4 (50%)	4 (44.4%)
Spiritual but not religious	3 (37.5%)	5 (55.5%)
Religious but not spiritual	---	---
Neither	1 (12.5%)	---
Ethnic Background		
White/Caucasian/European	6 (75.0%)	7 (77.8%)
East Asian	1 (12.5%)	1 (11.1%)
South Asian	1 (12.5%)	---
Undisclosed	---	1 (11.1%)
Country of Birth		
Canada	---	8 (88.8%)
USA	3 (37.5%)	1 (11.1%)
India	1 (12.5%)	---
New Zealand	1 (12.5%)	---
Ireland	1 (12.5%)	---
Scotland	1 (12.5%)	---

Philippines	1 (12.5%)	---

3.2 Delphi Results Round 1:

3.2.1 Assessment of Item Relevancy and Representativeness

In the first round of review, the 109-item pool achieved an overall Scale Content Validity Index (S-CVI) of 72% (I-CVI range: 0% - 100%) and 80% (I-CVI range: 22% - 100%) by SMEs and PAG, respectively (Table 2). Only 2 domains, Seeking to Understand and Global Compassion achieved a S-CVI/Ave greater than 80% among SMEs in Round 1 (Table 2). Overall, reviewers felt that the items collectively were either quite or highly representative of their respective domains, indicated by a median representative rating of 3 or 4 (Appendix C). After reviewing the SME and PAG feedback and cross-checking against other domain items, 8 potential additional items were flagged for further discussion by our team (Appendix D).

Table 2. SME and PAG Consensus for Item Relevancy: Scale-Content Validity Index from Delphi Round 1 and Round 2.

	SME Round 1 (n=6)	**SME Round 2 (n=8)**	**PAG Round 1 (n=9)**	**PAG Round 2 (n=9)**
	S-CVI/Average (I-CVI Range)			
Overall Compassion Measure (Round 1: 109 items) (Round 2: 84 items)	72% (0%-100%)	84% (50%-100%)	80% (22%-100%)	86% (44%-100%)
Domain 1 (Virtuous Response) (Round 1: 22 items) (Round 2: 19 items)	75% (0%-100%)	83% (63%-100%)	77% (22%-100%)	84% (44%-100%)
Domain 2 (Relational Space) (Round 1: 10 items) (Round 2: 7 items)	52% (17%-100%)	77% (50%-100%)	74% (44%-100%)	83% (67%-100%)
Domain 3 (Seeking to Understand) (Round 1: 25 items) (Round 2: 17 items)	83% (50%-100%)	90% (71%-100%)	76% (50%-100%)	93% (67%-100%)
Domain 4 (Relational Communication) (Round 1: 26 items) (Round 2: 22 items)	76% (50%-100%)	90% (75%-100%)	83% (56%-100%)	86% (63%-100%)

Domain 5 (Attending to Needs) **(Round 1: 22 items)** **(Round 2: 16 items)**	59% (33%-83%)	74% (50%-100%)	86% (67%-100%)	81% (56%-100%)
Domain 6 (Global Compassion) **(Round 1: 4 items)** **(Round 2: 3 items)**	87% (67%-100%)	92% (88%-100%)	78% (44%-89%)	89% (No range)

Note:
SME = Subject Matter Expert
PAG = Patient Advisory Group
S-CVI/Average = Scale-Content Validity Index/Average
I-CVI = Item-Content Validity Index

After analyzing the results of round 1 feedback, a total of 28 items were discarded (SME I-CVI range: 0% - 83%), 10 were re-worded (SME I-CVI range: 33% - 83%), 3 added, and 72 retained unmodified for round 2 (SME I-CVI range: 33% - 100%) (Figure 2; Appendix C and D). Items that received a large I-CVI discrepancy between SMEs and the PAG were either modified or retained and flagged for further testing during the cognitive interview stage. A sample of items that were either discarded, modified, or added are provided in Appendix D.

Figure 2. Flow diagram of item pool at each phase of review.

3.2.2 Assessment of the Recall Period

In terms of the 7-day recall period contained within the question stem, the SMEs had a median agreement rating of 5 (range: 4 – 5). The PAG provided a median agreement of 4, with a large variance in their responses (range: 2 – 5). Four participants were unsure why the recall period was restricted to the last 7 days, as patients living with an incurable, life-limiting illness likely had lengthier interactions with the healthcare system. Given the purpose of the measure and high median agreement ratings, however, the team felt that the recall period was sufficiently justified.

3.2.3 Assessment of the Relevancy of the Response Scale

In appraising the *relevancy of the response scale to measuring patient's experience of compassion,* SMEs had a median rating of 4 (range: 2 – 4) and the PAG had a median rating of 4 (range: 3 – 4). As 3 SMEs rated the relevancy of this frequency response scale ≤ 3 and queried the appropriateness of this type of scale (never - sometimes – usually – always) in measuring compassion, the response scale was changed to a 5-point Likert scale of agreement (strongly disagree – disagree – neutral – agree – strongly agree). The rationale for this decision was based on SME feedback that a frequency scale rating would vary across

settings (e.g. acute palliative care vs. hospice), is affected by patient-to-staff ratios, and is not an indicator of whether patients actually experienced compassion. For example, while patients in an acute palliative care setting may recall a significant number of compassionate behaviours over the course of 7 days, resulting in a high score in the compassion measure, they may not personally experience these specific behaviours as compassionate. Rather, both SMEs and our previous research suggest that it is the quality of these experiences, not the quantity of these experiences, that is the most salient indicator of compassion [14]. The question stems were also modified based on SME and PAG feedback to reflect a group of HCPs being assessed as opposed to a single HCP being evaluated. Evaluating the care team was congruent with the model of care in our target settings, and the sentiment among SMEs and the research team that an uncompassionate experience with a single HCP could outweigh the positive experiences of compassion that patients experienced from the larger care team — i.e. the summation of their interactions with their care team versus the experience of one clinician was considered a more accurate indicator of compassion within our patient populations.

3.3 Delphi Results Round 2:

3.3.1 Assessment of Item Relevancy and Representativeness

In the second round of review, the entire draft compassion measure (84 items) achieved our targeted Content Validity Index (S-CVI/Ave) of 80% [35], with an S-CVI of 84% (I-CVI range: 50% - 100%) among SMEs and 86% (I-CVI range: 44% - 100%) within the PAG (Table 1). The S-CVI/Ave within each domain improved between rounds to >80% (S-CVI range: 81% - 89%) among the PAG and ≥74% (S-CVI range: 74% - 92%) among SMEs. The number of measure items was reduced based on participants' preferences for alternatively worded items: 16 items were discarded due to either a low I-CVI or an alternative item preference, 1 was re-worded, and 67 items were retained unmodified (Figure 2). The 10 items that were re-worded after round 1 all achieved higher I-CVIs by the SMEs in round 2, indicating content improvement. The 3 items that were added (Appendix D) achieved an I-CVI between 75% and 88% among SMEs, and 100% among PAG members, and were therefore included in cognitive interviews.

Both the SME and PAG panels indicated that the items were quite representative or highly representative of their respective domains, with median ratings of either 3 or 4 (Appendix C). As all domains achieved a median representativeness rating of at least 3 (quite representative) in comparison to some of the

domains receiving a rating of 2 (somewhat representative) in the 1st round, this indicated improvement with respect to content coverage (Appendix C). For the SME and PAG panels, the median relevancy rating of the new agreement response scale was 4 (no range) and 4 (range: 3 – 4) respectively. As the relevancy *range* of this new response scale to measuring patients' experience of compassion decreased from the previous round (i.e. no variance in ratings within the SMEs), this indicates an improvement with respect to this aspect of the measure. Lastly, because the proposed recall period was unchanged after round 1 feedback, we refrained from duplicating this survey question in the second round of review. Finally, the additional PAG question, assessing the importance of a patient-reported measure of compassion to inform patient care, resulted in a median importance rating of 5 (range: 4 - 5).

3.3.2 Assessment of Clinical Sensibility: Item clarity and cognitive understanding via cognitive interviews

Sixteen patients from acute palliative care (n=4), palliative home care (n=4), long term care (n=4), and hospice care (n=4) participated in the cognitive interviews (Table 3), until saturation was reached, resulting in an additional 14 items being discarded (Figure 2). Five items were discarded due to issues of comprehension, recall, and judgment. Nine items were discarded due to preference for alternative wording of a parallel item. Finally, one of the retained items was reworded to improve clarity. Overall, when asked how easy or difficult they felt it would be for a patient to answer the questions in thinking of their experiences with their Healthcare Providers over the past 7 days, the majority of the cognitive interview participants felt that the recall period was appropriate.

Table 3. Demographics: Cognitive Interview Patient Participants (n=16).

	Patient Participants (n=16)
	Frequency (Percentage)
Care Location	
Acute care	4 (25.0%)
Home care	4 (25.0%)
Residential care	4 (25.0%)
Hospice Care	4 (25.0%)
Highest, Completed Level of Education	
No Formal Education	0 (0.00%)
Some Elementary	0 (0.00%)
Elementary completed	0 (0.00%)
Some High school	2 (12.5%)
High School- Grade 12 completed	5 (31.3%)
Some University or Technical school	2 (12.5%)
University/College/Technical school completed	4 (25.0%)

Post graduate University completed	3 (18.8%)
Gender	
Male	10 (62.5%)
Female	6 (37.5%)
Median Age (Range) (n=16)	61 (37 – 92)
Religious groups affiliation	
Roman Catholic	4 (25.0%)
None	3 (18.8%)
Other (Mormon; Muslim (n=2); Anglican; Presbyterian; Church of Jesus Christ; Spiritualist)	7 (43.8%)
Jewish	1 (6.25%)
Jehovah's Witness	1 (6.25%)
Religiousness or Spirituality	
Spiritual and religious	10 (62.5%)
Spiritual but not religious	5 (31.3%)
Religious but not spiritual	0 (0.00%)
Neither	1 (6.25%)
Ethnic Background	
White/Caucasian/European	11 (68.8%)
South Asian (e.g. East Indian, Pakistani, Sri Lankan)	3 (18.8%)
East Asian (e.g. Chinese, Korean, Vietnamese)	1 (6.25%)
Aboriginal Peoples of Canada	1 (6.25%)
Reason for Care	
Pneumonia; Bowel obstruction surgery; Dementia	1 (6.25%)
Chronic Mental Health/Depression	1 (6.25%)
Chronic COPD	3 (18.8%)
Chronic Arthritis	1 (6.25%)
Advanced Cancer	9 (56.3%)
Advanced Liver Disease	1 (6.25%)
Marital Status	
Married	6 (37.5%)
Single	6 (37.5%)
Separated	1 (6.25%)
Divorced	1 (6.25%)
Widow(er)	2 (12.5%)

4 Discussion

This study established content-related validity evidence for a newly developed compassion measure to be utilized by clinicians and researchers, in order to enhance an aspect of quality care, that while being identified as imperative to patients, is increasingly lacking from their experience of healthcare. In this study, the two Delphi rounds revealed that both the PAG and SMEs felt the items and domains within the draft compassion measure were

relevant, representative, and collectively provided theoretical coverage to the overarching construct of interest. Cognitive interviews resulted in the measure being reduced to 54 items (Figure 2).

The Delphi results came from not only SMEs on the topic of compassion, but also from a panel of Patient Advisors. This additional, intentional and strategic assessment of content validation was congruent with the tenets of patient-reported measure development, while also addressing a significant limitation of existing measures of compassion—the need to integrate the patient perspective across all stages of measure development and validation [14].

In instrument development, incongruence between the item content (individual questions reflective of the construct of interest) of the construct of interest (the overarching phenomenon of interest) and content domain (a collection of questions comprising a major component of the construct of interest) of a measure has been identified as a common and recurring problem [15, 37, 39-41]. In an effort to mitigate any confusion amongst the PAG and SMEs, our team was vigilant in ensuring that clear and concise instructions were provided to our reviewers, along with definitions for each domain and the overarching construct, as depicted within the Patient Compassion Model [4]. This allowed the experts to compare each item against its respective domain definition and where necessary, relevancy to the target population. The PAG provided additional insight, which informed the modification of a number of items that the SMEs had not suggested. As such, we argue that patients should not only be considered in the development of patient-reported measures, but are integral, including the construction and assessment of the measure across all stages of development.

As noted in a recent review, existing measures of compassion have significant limitations [13]. A fundamental issue within most of these measures is a lack of validity evidence based on content and clinical sensibility testing. Therefore, it is uncertain whether existing compassion measures are in fact measuring compassion or some other related care construct. In the 3 studies [42-44] that reported information on content validation, the response rates of SMEs for these 3 studies were undisclosed, making it difficult to deduce an expected response rate for our current study. In using specific inclusion and exclusion criteria for selecting the experts for the present study, we ensured that we received feedback from SMEs on the topic of compassion specifically, as opposed to including experts on related constructs of empathy or person-centred care. Owing to this judicious process, we feel confident that our results are a valid depiction of the compassion construct as determined by a diverse group of participants inclusive of academic and patient experts.

Clinical sensibility testing with a measures' target population, utilizing cognitive interviews, is an important component of establishing the construct validity of a measure. Unfortunately, this important step occurs infrequently or is under-reported in the development of compassion measures and patient-reported outcome measures in general [13, 45]. As a result, it is difficult to discern whether these measures rest on a valid content evidence base, and are relevant and comprehendible to end users. Patients' views regarding the ease of answering the items and their ability to rate personal experiences with their Healthcare Providers within the 7-day recall period suggest that the proposed measure is appropriate for our target patient populations and settings.

Validity is established through an ongoing process of accumulating and evaluating evidence of multiple types and sources in order to support the proposed interpretation and uses of an instrument's scores for a given purpose or purposes [46]. While this study has provided content-related validity evidence for our compassion measure, subsequent stages of validation and reliability assessments will provide additional evidence based on the measure's internal structure (through exploratory and confirmatory factor analysis) and relationships to other variables (convergent and discriminant validity evidence). A valid and reliable tool to measure compassion not only advances research on compassionate care, but provides healthcare providers with a tool to measure patients' experiences of compassion on a routine basis, allowing them to modify caregiving on an individual, team, and organizational level accordingly.

5 Conclusions

Use of a modified Delphi technique and cognitive interviews helped us to systematically reduce the initial item pool of the compassion measure, establishing initial content-related validity evidence in the process. Additional testing of the psychometric properties of the patient-reported compassion measure is needed to further establish validity and ensure that the final measure accurately and consistently measures what it purports to.

Declarations

***Authorship*:** All authors 1) made substantial contributions to the conception and design of the study, acquisition of data, and analysis and interpretation of data; 2) were involved in drafting the manuscript or revising it critically for important intellectual content; 3) gave final approval of the version to be published and have participated sufficiently in the work to take public responsibility for appropriate portions of the content; and 4) agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding: This study was funded through a Canadian Institutes of Health Research Project Scheme Grant (#364041).

Competing interests: The authors declare that they have no competing interests.

Ethics approval and consent to participate: This study was approved by the University of Calgary Conjoint Health Research Ethics Board (REB #17-0754).

Availability of data and material: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Acknowledgements: The authors would like to acknowledge the following individuals for their support in this study:

Subject Matter Experts:

Dr. Mary Vachon, Dr. Nathan Consedine, Dr. Philip Larkin, Dr. Betty Ferrell, Dr. Stephen Post, Dr. Stephen Smith, Dr. Latha Chandran, Dr. Antonio Fernando;

Mark Hall, Research Associate, Faculty of Nursing, University of Calgary: for providing support with the data analysis.

References

1. Weldring T, Smith S. Patient-reported outcomes (PROs) and Patient-reported Outcome Measures (PROMs). Health Services Insights. 2013;6 61-68.
2. Lakhani A. Indicators for measuring patient experience. 2013. https://www.patientlibrary.net/tempgen/105346.pdf. Accessed December 18, 2018.
3. Kingsley C, Patel S. Patient-reported outcome measures and patient-reported experience measures, BJA Education. 2017 14(4):137-144.
4. Sinclair, S., McClement, S., Raffin-Bouchal, S., et al. Compassion in health care: an empirical model. J Pain Symptom Manage. 2016 51: 193–203.
5. Francis R. Report of the Mid Staffordshire NHS Foundation Trust public inquiry. London: The Stationary Office. 2013.
6. Heyland D, Dodek P, Rocker G, Canadian Researchers End-of-Life Network (CARENET)., et al. What matters most in end-of- life care: perceptions of seriously ill patients and their family members. CMAJ. 2006 174(5):627–33.

7. Crowther J, Wilson K, Horton S, et al. Compassion in healthcare: lessons from a qualitative study of the end of life care of people with dementia. J R Soc Med. 2013 106(12):492–7.
8. Lown B, Rosen J, Marttila, J. An agenda for improving compassionate care: a survey shows about half of patients say such care is missing. Health Aff (Millwood). 2011 30:1772–8.
9. Willis L. Raising the bar: the shape of caring review. London: Health Education England. 2015.
10. Maclean R. The Vale of Leven Hospital Inquiry. Edinburgh: APS Group. 2014.
11. National Health Service Commissioning Board Chief Nursing Officer and Department of Health Chief Nursing Adviser. Compassion in practice: nursing, midwifery and care staff. Our vision and strategy. Leeds: Department of Health. 2012.
12. Reader T, Gillespie A. Patient neglect in healthcare institutions: a systematic review and conceptual model. BMC Health Serv Res. 2013 13:156–70.
13. Sinclair S, Russell LB, Hack TF, et al. Measuring compassion in healthcare: A comprehensive and critical review. Patient. 2016 10(4):389-405.
14. Sinclair S, Beamer K, Hack TF, et al. Sympathy, empathy, and compassion: a grounded theory study of palliative care patients' understandings, experiences, and preferences. Palliat Med. 2016. doi:10.1177/0269216316663499.
15. Sinclair S, Jaggi P, Hack TF, et al. A practical guide for item generation in measure development: Insights from the development of a patient-reported experience measure of compassion. Journal of Nursing Measurement. 2019. *In Press*.
16. DeVellis R. Scale development: Theory and applications. (2nd edn.). Sage Publications, Newbury Park, CA. 2003.
17. Saris W. Design, evaluation, and analysis of questionnaires for survey research. Wiley. 2014.
18. Hinkin T. A brief tutorial on the development of measures for use in survey questionnaires. Sage Publications. 1998 Vol 1:1; 104-121.
19. Sinclair S, Jaggi P, Hack TF, et al. Assessing the credibility and transferability of the patient compassion model in non-cancer palliative populations. BMC Palliative Care. 2018 17:108.
20. Lynn M. Determination and quantification of content validity. Nursing Research. 1986 35(6), 382-385.
21. Yaghmale F. Content validity and its estimation. Journal of Medical Education. 2003 3(1); 25-27.

22. Messick S. Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–103). New York, NY: American Council on Education & Macmillan Publishing Company. 1989.
23. Sireci SG. The construct of content validity. Social Indicators Research, 1998 45(1-3), 83–117. doi:10.1023/A:1006985528729.
24. Burns K, Duffett M., Kho M, et al. A guide for the design and conduct of self-administered surveys for clinicians. CMAJ. 2008 179(3).
25. Hsu C and Sanford B. The delphi technique: Making sense of consensus. 2007 12(10), 1.
26. Jones J and Hunter D. Consensus methods for medical and health services research. Br Med J. 1995 311(7001):376–80.
27. Sinclair S, Norris JM, McConnell SJ, et al. Compassion: a scoping review of the healthcare literature. J Pain Symptom Manage BMC Palliat Care. 2016 15:6–20.
28. Canadian Institutes of Health Research. SPOR Support Units http://www.cihr-irsc.gc.ca/e/45859.html. Accessed March 11th, 2019.
29. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. Journal of Biomedical Informatics, 2009 42(2), 377-381.
30. Slocumb E and Cole F. A practical approach to content validation. Applied Nursing Research. 1991 4, 192–195.
31. Grant J and Davis L. Selection and use of content experts for instrument development. Research in Nursing & Health. 1997 20, 269–274.
32. Willis GB. Cognitive Interviewing: A tool for improving questionnaire design. Thousand Oaks, CA: Sage. 2005.
33. Polit D and Beck C. The content validity index: Are you sure you know what's being reported? Critique and recommendations. Research in Nursing & Health. 2006 29:489-497.
34. Polit D, Beck C, and Owen S. Is the CVI and acceptable indicator of content validity? Appraisal and recommendations. Research in Nursing and Health. 2007 30:459-467.
35. Davis, L.L. (1992). Instrument review: Getting the most from a panel of experts. Applied Nursing Research, 5, 194–197.

36. Ritchie J, Spencer L, O'Connor W. Carrying out Qualitative Analysis. Qualitative Research Practice. London, Sage. 2003 219-262.
37. Grant J and Davis L. Selection and use of content experts for instrument development. Research in Nursing & Health. 1997 20, 269–274.
38. Sinclair S, Beamer K, Hack TF, et al. Sympathy, empathy, and compassion: a grounded theory study of palliative care patients' understandings, experiences, and preferences. Palliat Med. 2016. doi:10.1177/0269216316663499.
39. Rattray J and Jones M. Essential elements of questionnaire design and development. Journal of Clinical Nursing. 2007. Blackwell Publishing Ltd.
40. Nichols P and Sugrue B. The lack of fidelity between cognitively complex constructs and conventional test development practice. Educational Measurement: Issues and Practice. 1999.
41. Bowling, A. Research Methods in Health. Open University Press, Buckingham. 1997.
42. Lown BA, Muncer SJ, Chadwisk R. Can compassionate healthcare be measured? The Schwartz Center Compassionate Care ScaleTM. Patient Educ Couns. 2015 98(8):1005–10.
43. Lee Y, Seomun G. Development and validation of an instrument to measure nurses' compassion competence. Appl Nurs Res. 2016 30:76–82.
44. Burnell L, Agan D. Compassionate care can it be defined and measured? The development of the compassionate care assessment tool. Int J Caring Sci. 2013 6:180–7.
45. Collins D. Analysing and interpreting cognitive interview data: A qualitative approach. Quest 2007 Proceedings. 2007 64-73. https://wwwn.cdc.gov/QBank/Quest/2007/QUEST%202007%20Proceedings-all%20papers.pdf#page=74 Accessed on March 11th, 2019.
46. American Educational Research Association, American Psychological Association, National Council on Measurement in Education, & Joint Committee on Standards for Educational and Psychological Testing (U.S.). Standards for educational and psychological testing. Washington, DC: American Educational Research Association. 2014.

Supporting Information

<u>Appendix A</u>. Cognitive Interviews: Sample Questions.

Appendix B. Independent Reviewer Decision-Making Guide for Item Modification, Additions, or Deletions.

Appendix C. Number of item deletions, modifications, and additions, and final number of items per Delphi round.

Appendix D. Sample of items that were discarded, re-worded, or retained in both rounds of SME and PAG review.