Measure Information

This document contains the information submitted by measure developers/stewards, but is organized according to NQF's measure evaluation criteria and process. The item numbers refer to those in the submission form but may be in a slightly different order here. In general, the item numbers also reference the related criteria (e.g., item 1b.1 relates to sub criterion 1b).

Brief Measure Information

NQF #: 3726

Corresponding Measures:

Measure Title: Serious Illness Survey for Home-Based Programs

Measure Steward: RAND Corporation

sp.02. Brief Description of Measure:

The proposed measures are derived from the Serious Illness Survey for Home-Based Programs, a 36-item questionnaire designed to measure the care experiences of patients receiving care from home-based serious illness programs. Home-based serious illness programs provide care for seriously ill patients at their private residences (i.e., in their homes or assisted living facilities, not in institutions like skilled nursing facilities). Programs are staffed by interdisciplinary teams that provide support for palliation of symptoms, assist with coordination of care, answer questions after-hours, provide medication management, and assist with advance care planning (Cohn et al., 2017). Teams consist of clinicians (e.g. physicians, nurse practitioners) that oversee care, as well as clinical and supportive staff that make home visits (e.g. registered nurses, social workers, CNAs). Programs serve patients with a life expectancy that ranges from 1-5 years and have enrollment criteria based on diagnosis, symptom burden, functional status, and/or prior health care utilization.

The five proposed multi-item measures are:

1. Communication
2. Care Coordination
3. Help for Symptoms
4. Planning for Care
5. Support for Family and Friends

The two proposed single-item measures are:

1. Overall Rating of the Program
2. Willingness to Recommend the Program

Appendix A presents the survey items included in each measure, including response options for each item. Measure scores are "top-box" scores that reflect the percent of respondents who select the most positive response category(ies) in response to the survey item(s) within the measure.

Citation:

Cohn J, Corrigan J, Lynn J, Meier D, et al. Community-Based Models of Care Delivery for People with Serious Illness. National Academy of Medicine Discussion Paper. Available at https://nam.edu/wp-content/uploads/2017/04/Community-Based-Models-of-Care-Delivery-for-People-with-Serious-Illness.pdf.

1b.01. Developer Rationale:

Caring for high-need, high-cost seriously ill people is an urgent priority because they account for an increasing proportion of users of the health care system, their care is expensive, and they are more likely than others to be affected by variations in health care quality and safety due to their frequent use of care (Blumenthal, Chernof et al. 2016).

To address the health care needs of seriously ill people, there has been rapid growth of home-based serious illness programs (Long, Abrams et al., 2017; Cohn, Corrigan et al., 2017; NASEM, 2018; Olson et al., 2019). Although measuring, monitoring, and incentivizing high quality care is critical for this vulnerable population, to date, no standardized measures have been developed for programs that provide care to seriously ill people in their homes. Experts have highlighted the need for standardized measures of patient and family care in such programs (NQF, 2020). Quality measures are critical for the determination of value in both current payment models, such as Medicare Advantage and Medicare accountable care organizations (ACOs), and future alternative payment models under consideration to serve seriously ill populations.

Patient- and family-centeredness is particularly important for seriously ill patients, given great variability across patients with regard to both preferences for care intensity and tradeoffs between quality and length of life. Surveys of patients and their family caregivers are the main means of assessing patient- and family-centeredness of care, and survey results can be used to identify areas of patient and family experiences of care that need improvement, monitor quality over time, and be incorporated into value-based models to allow for benchmarking and comparison of programs (Elliott, Cohea, et al., 2015; Rodriguez, von Glahn, et al., 2009; Quigley and McCleskey, 2020).

Expert consensus has identified core areas of high-quality serious illness care about which patients and families are the best and only source of information, including person- and family-centered communication, access and responsiveness, shared decision making and care planning in support of patient goals, care coordination (including medication management), symptom management, and attention to caregiver needs and social determinants of health (Long, Abrams et al., 2017; Cohn, Corrigan et al., 2017). In prior studies, seriously ill patients and/or their family caregivers have reported substantial unmet need in each of these areas, including inadequate communication and reports of not feeling “heard or understood” by their health care providers (Ahluwalia et al., 2022a; Casarett et al., 2008; Covinsky et al., 2000; Dy et al., 2008; Teno et al., 2004), not getting help as soon as needed or on evenings and weekends (Anhang Price et al., 2018), receiving care that is not consistent with their preferences (Khandelwal et al., 2017; Teno et al., 2015), and not receiving desired help with pain and other symptoms (Ahluwalia et al., 2022b; Anhang Price et al., 2018; Teno et al., 2004; Teno et al., 2015).

Citations:

Ahluwalia SC, Vegetabile BG, Edelen MO, Setodji CM, Rodriguez A, et al. (2022a). MACRA Palliative Care Quality Measure Development – Testing Summary Report. Measure Name: Feeling Heard and Understood. *RAND Health Q.* 9(3): 3.

Ahluwalia SC, Vegetabile BG, Edelen MO, Setodji CM, Rodriguez A, et al. (2022b). MACRA Palliative Care Quality Measure Development – Testing Summary Report. Measure Name: Receiving Desired Help for Pain. *RAND Health Q.* 9(3): 3.

Anhang Price R, Stucky B, Parast L, Elliott MN, Haas A, Bradley M, Teno JM. (2018). Development of Valid and Reliable Measures of Patient and Family Experiences of Hospice Care for Public Reporting. *J Pall Med.* 21(7):924-932.

Blumenthal, D., B. Chernof, T. Fulmer, J. Lumpkin and J. Selberg (2016). "Caring for High-Need, High-Cost Patients - An Urgent Priority." *N Engl J Med.* 375(10): 909-911.

Casarett, D., Pickard, A., Bailey, F. A., Ritchie, C. S., Furman, C. D., Rosenfeld, K., Shreve, S., & Shea, J. (2008). A nationwide VA palliative care quality measure: the family assessment of treatment at the end of life. *J Palliat Med.* 11(1), 68-75.

Cohn J, Corrigan J, Lynn J, et al. (2017). Community-Based Models 20

Covinsky, K. E., Fuller, J. D., Yaffe, K., Johnston, C. B., Hamel, M. B., Lynn, J., Teno, J. M., & Phillips, R. S. (2000). Communication and decision-making in seriously ill patients: findings of the SUPPORT project. The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. *J Am Geriatr Soc.* 48(5 Suppl), S187-193.

Dy, S. M., Shugarman, L. R., Lorenz, K. A., Mularski, R. A., & Lynn, J. (2008). A systematic review of satisfaction with care at the end of life. *J Am Geriatr Soc.* 56(1), 124-129.

Elliott MN, Cohea CW, Lehrman WG, et al. (2015). Accelerating improvement and narrowing gaps: Trends in patients' experiences with hospital care reflected in HCAHPS public reporting. *Health Serv Res.* 50(6):1850-1867. 10.1111/1475-6773.12305.

Khandelwal, N., J. R. Curtis, V. A. Freedman, J. D. Kasper, P. Gozalo, R. A. Engelberg and J. M. Teno. (2017). "How Often Is End-of-Life Care in the United States Inconsistent with Patients' Goals of Care?" *J Palliat Med.* 20(12): 1400-1404.

Long P, Abrams M, Milstein A, et al., eds. (2017). Effective Care for High-Need Patients: Opportunities for Improving Outcomes, Value, and Health. Washington, DC: National Academy of Medicine.

National Academies of Sciences, Engineering, and Medicine. (2018). Models and Strategies to Integrate Palliative Care Principles into Care for People with Serious Illness: Proceedings of a Workshop. Washington, DC: The National Academies Press.

National Quality Forum (NQF). (2020). Issue Brief: Opportunities for Advancing Quality Measurement in Community-Based Serious Illness Care. Available at: https://store.qualityforum.org/products/issue-brief-opportunities-or-advancing-quality-measurement-in-community-based-serious-illness-care.

Olson A, Harker M, Saunders R, et al. (2019) Innovations in Medicare Advantage To Improve Care for Seriously Ill Patients: Duke Margolis Center for Health Policy, Robert Wood Johnson Foundation.

Quigley DD, McCleskey SG. (2020). Improving care experiences for patients and caregivers at end of life: A systematic review. *Am J Hosp Palliat Care*. 38(1):84-93.

Rodriguez HP, von Glahn T, Elliott MN, et al. (2009). The effect of performance-based financial incentives on improving patient care experiences: A statewide evaluation. *J Gen Intern Med*. 24(12):1281-1288.

Teno, J. M., Clarridge, B. R., Casey, V., Welch, L. C., Wetle, T., Shield, R., & Mor, V. (2004). Family perspectives on end-of-life care at the last place of care. *JAMA*, 291(1), 88-93.

Teno, J. M., Freedman, V. A., Kasper, J. D., Gozalo, P., & Mor, V. (2015). Is care for the dying improving in the United States? *J Palliat Med.* 18(8), 662-666.

---

sp.12. Numerator Statement: Measure scores are "top-box" scores that reflect the percent of respondents who select the most positive response category(ies) in response to the survey item(s) within the measure. Therefore, the numerator is the number of respondents who select the most positive response category(ies) in response to the survey items within the measure.

sp.14. Denominator Statement: Survey respondents are patients receiving care from home-based serious illness programs. Survey eligibility criteria and exclusions are detailed below in sections sp.16 – sp.18. Screener questions and tailored non-applicable response options (e.g., I did not want help for my pain) are used to identify respondents who are and are not eligible to respond to survey items included in evaluative measures. Therefore, denominators vary by survey item (and corresponding multi-item measures, if applicable) according to the eligibility of respondents for each item.

sp.16. Denominator Exclusions:

The Serious Illness Survey for Home-Based Programs is designed for administration to adult patients who are currently enrolled in home-based serious illness programs. Patients are excluded from the survey sample if they:

- Are under age 18
- Receive care from a serious illness program in a setting OTHER than home or an assisted living facility (e.g., in a nursing home or other long-term care facility)
- Are known to have been discharged to hospice
- Are known to have died
- Have been enrolled in the serious illness program for less than six weeks as of the date of survey sampling

In keeping with the Medicare CAHPS Survey (https://www.cms.gov/files/document/ma-pdp-cahps-qapts-v11-complete-manual.pdf), a survey is considered partially completed if there are responses to at least one measure and for less than 50 percent of survey items that are applicable to all. A survey is considered completed if there are responses to at least one measure and for 50 percent or more of the survey items that are applicable to all. Final analytic datasets include all completed and partially completed surveys.

There are no explicit exclusions based on language; the survey is available in English and Spanish.

Measure Type: Outcome: PRO-PM

sp.28. Data Source:

Instrument-Based Data

sp.07. Level of Analysis:

Other

IF Endorsement Maintenance – Original Endorsement Date:

Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

sp.03. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results?:

1. Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria

---

Please separate added or updated information from the most recent measure evaluation within each question response in the Importance to Measure and Report: Evidence section. For example:

Current Submission:

Updated evidence information here.

Previous (Year) Submission:

Evidence from the previous submission here.

1a.01. Provide a logic model.

Briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

[Response Begins]

Home-based serious illness programs provide care for people with advanced, progressive illnesses in their private residences (i.e., in their homes or assisted living facilities, not in institutions like skilled nursing facilities). Programs are staffed by interdisciplinary teams that provide support for palliation of symptoms, assist with coordination of care, answer questions after-hours, provide medication management, and assist with advance care planning (Cohn et al., 2017). Teams consist of health care professionals that may include physicians, nurse practitioners, physician assistants, nurses, social workers, spiritual counselors, bereavement staff, and CNAs.

We identified guiding principles, features and competencies of high-quality, home-based serious illness care by synthesizing recommendations from a Roundtable on Quality Care for People with Serious Illness at the National Academies of Sciences, Engineering, and Medicine (Long, Abrams et al., 2017; Cohn, Corrigan et al., 2017), seeking feedback from a panel of technical experts, and conducting qualitative interviews with home-based program health care providers, and patients and family caregivers receiving care from these programs. Table 1a.01 summarizes these core aspects of high-quality home-based serious illness care, and the key processes of care required to achieve them.

Table 1a.01. Core Aspects of High-Quality Home-Based Serious Illness Care.

Aspect of High-Quality Serious Illness Care	Description and Key Processes of Care

Person-/Family-Centered	• Care that is "grounded in mutually beneficial partnerships among health care providers, patients, and families…" in which seriously ill people and families "…determine how they will participate in care and decision-making" throughout the course of care, including the end of life (adapted from Johnson B.H., Abraham M.R., 2012)
	• Includes:
	  ○ Open, clear and sensitive two-way communication
	  ○ Treating seriously ill person and family with dignity and respect
	  ○ Emotional and spiritual support
	  ○ Culturally competent care

Shared Decision Making and Advance Care Planning in Support of Patient Goals	• Understand seriously ill person's and family's desired role in decision making, and involve them in that manner • Understand what is important to the seriously ill person and family • Provide desired information regarding disease trajectory, prognosis, treatment options ○ Contextualized in context of values and preferences • Elicit and help, if needed, to formulate preferences for care now and in the future • Instill trust that program will honor preferences • Develop care plan to honor preference, adapt it over time
Care Continuity and Coordination, including Medication Management	• Deliver care using multidisciplinary care teams with experienced health care professional acting as the coordinator and communication hub • Ensure that seriously ill person and family have a trusted point of contact • Build processes to promote appropriate and informed transitions among providers within the care team and across care settings • Carefully manage and reconcile medications
Access and Responsiveness	• Provide access to medical care that is consistent with seriously ill person's and family preferences and current medical knowledge regarding effectiveness • Timely provider responsiveness • 24/7 availability
Symptom Management / Palliation	• Recognize, assess, and understand the cause of symptoms, including physical, emotional, social, or spiritual sources • Achieve seriously ill person's desired level of control for each symptom • Manage using interdisciplinary approach that uses both pharmacologic and non-pharmacologic approaches
Attention to Caregiver Needs	• Identify needs and concerns of family caregivers • Provide needed information and training to safely care for the seriously ill person

Aspect of High-Quality Serious Illness Care	Description and Key Processes of Care
Attention to Social Determinants of Health	• Understand impacts of social determinants on seriously ill person and their family • Address through direct provision of services, or link/refer seriously ill person and family to available social services

Table 1a.01 describes core aspects of high-quality home-based serious illness care, and the key process of care that achieve them, including person- and family-centered care, shared decision making and advance care planning in support of patient goals, care continuity and coordination, including medication management, access and responsiveness, symptom management, attention to caregiver needs, and attention to social determinants of health.

The Serious Illness Survey for Home-Based Programs assesses the degree to which each of these core aspects of care are being delivered by asking patients to describe experiences of care that result from each aspect. Table 1a.02 notes the structure (provision of care by a home-based program), processes, and measured expected outcomes.

Table 1a.02. Logic Model for Serious Illness Survey for Home-Based Programs Measures.

STRUCTURE	PROCESSES	OUTCOMES
Program, staffed by an interdisciplinary team, that visits seriously ill patients at home.	Regular, person-and family-centered communication between patient, family, and interdisciplinary home-based program team • Review care provided by other providers, all medications, and whether help is needed to participate in everyday activities • Understand desired role in decision-making, involve patient and family in that manner • Understand what is important to patient and family, incorporate into care plan Answer calls after hours and between visits to address questions and schedule additional visits, as needed Screen, diagnose, treat, and routinely monitor symptoms	**Communication** Patient reports that the program: • Spent enough time • Explained things in a way they could understand • Listened carefully • Cared about them as a whole person • Made them feel "heard and understood" **Care Coordination** Patient reports that the program: • Knew important information about their medical history • Talked with them about care or treatment from other providers, all medications they were taking, and how to get help with everyday activities • Provided help when they needed in between home visits **Help for Symptoms** Patient reports that the program provided as much help as wanted for pain, trouble breathing, and feelings of anxiety or sadness **Care Planning** Patient reports that the program talked with them about: • What to do during a health emergency • What is important in their life • What their health care options would be if they got sicker **Support for Family and Friends** Patient reports that the program: • Involved their family members and friends in discussions as much as they wanted • Provided as much emotional support as family and friends wanted

Table 1a.02, the Logic Model for Serious Illness Survey for Home-Based Programs Measures, displays the Structure (the program, staffed by an interdisciplinary team, that visits seriously ill patients at home), the key processes of care provided by that program (regular person- and family-centered communication, answering calls after hours and between visits, and screening, diagnosing, treating, and routinely monitoring symptoms), and outcomes (communication, care coordination, help for symptoms, care planning, and support for family and friends).

Citations:

Cohn J, Corrigan J, Lynn J, et al.: Community-Based Models of Care Delivery for People with Serious Illness. NAM Perspectives, Discussion Paper. Washington, DC: National Academy of Medicine, 2017.

Johnson B.H., Abraham M.R. (2012). Partnering with Patients, Residents, and Families: A Resource for Leaders of Hospitals, Ambulatory Care Settings, and Long-Term Care Communities. Bethesda, MD: Institute for Patient- and Family-Centered Care.

Long P, Abrams M, Milstein A, et al., eds. (2017). Effective Care for High-Need Patients: Opportunities for Improving Outcomes, Value, and Health. Washington, DC: National Academy of Medicine.

National Academies of Sciences, Engineering, and Medicine: Models and Strategies to Integrate Palliative Care Principles into Care for People with Serious Illness: Proceedings of a Workshop. Washington, DC: The National Academies Press, 2018.

[Response Ends]

1a.02. Provide evidence that the target population values the measured outcome, process, or structure and finds it meaningful.

Describe how and from whom input was obtained.

[Response Begins]

In May 2017, serious illness care experts and stakeholders, including practicing palliative care physicians, patient advocates, academic researchers, quality and policy experts, government leaders, and health plan representatives, proposed a starter set of quality measures for assessing quality of home-based serious illness care; key among these were measures of patient care experiences (NQF, 2020).

To identify which aspect of patient care experience would be most meaningful to both home-based serious illness care providers and the seriously ill patients and families they serve, in 2018 and 2019, we conducted qualitative interviews with 16 patients, 16 family caregivers, five patients and caregivers together (i.e., joint interviews), and nine health care professionals (e.g., physicians, nurse practitioners, registered nurses, and social workers providing direct care to patients in home-based serious illness programs). Interviewees were recruited from a diverse set of home-based programs across regions of the United States. In all interviews, we discussed the range of services that home-based programs provide, and the aspects of care that patients and families most valued. In interviews with health care providers, we also inquired about the aspects of care most amenable to quality improvement.

Interviewees reported that they valued communicating with the members of the home program team, who they believed should be caring, supportive, and attentive to patients' and families' needs (as examined in the survey domains of Communication and Support for Family and Friends). They further noted the value of having the home program team be a "second set of eyes" regarding treatments, appointments, and medications received from multiple providers; appreciated being able to call someone from the program after hours to ask questions; and found it helpful when the program referred them to community resources, such as meal and grocery services, medical equipment chairs, and transportation services (examined in the Care Coordination domain). Interviewees also noted the programs' helpful role in providing symptom management that complemented symptom palliation offered by other primary health care providers, as much as desired by the patient (examined in Help for Symptoms domain), and in advance care planning; helping patients determine and express their care preferences, goals, and wishes to other providers; and promoting shared decision making (examined in the Care Planning domain).

Citation:

National Quality Forum (NQF). (2020). Issue Brief: Opportunities for Advancing Quality Measurement in Community-Based Serious Illness Care. Available at: https://store.qualityforum.org/products/issue-brief-opportunities-or-advancing-quality-measurement-in-community-based-serious-illness-care.

[Response Ends]

1a.03. Provide empirical data demonstrating the relationship between the outcome (or PRO) and at least one healthcare structure, process, intervention, or service.

[Response Begins]

Prior research has established that home-based serious illness care is associated with higher patient and family satisfaction with care (examined in the Overall Rating domain), as well as reduced symptom burden, particularly for depressive and anxious symptoms (Gomes et al., 2013; Singer et al., 2016; examined by the Help for Symptoms domain). There is also strong evidence that home-based serious illness care reduces hospital use (Singer et al., 2016) and increases the likelihood that patients die at home (Ahluwalia et al., 2018), the location preferred by most patients, perhaps indicating the role of home-based program teams in engaging patients and their families in discussions about wishes for care and what to do in the case of a health emergency (examined by the Care Planning domain).

As described in Section 1a.01, the core care process of home-based serious illness care is regular, person- and family-centered communication between the patient, family, and interdisciplinary home-based program team. A growing body of evidence supports the association between serious illness communication (examined in the Communication domain of the Serious Illness Survey for Home-Based Programs) and a range of outcomes. For example, effective serious illness communication has been shown to increase the likelihood that their wishes are known and followed (Detering et al., 2010; Mack et al., 2010), to improve overall patient and family experiences of care (Detering et al., 2010; Leung et al., 2012; Anhang Price et al., 2018), and to reduce health care spending (Wright et al., 2008; Zhang et al., 2009). Practical guidance for clinicians on how to initiate and conduct serious illness discussions is available to promote improved communication (Bernacki and Block, 2014).

Citations:

Ahluwalia, SC, Chen C, Raaen L, et al. (2018). A Systematic Review in Support of the National Consensus Project Clinical Practice Guidelines for Quality Palliative Care, Fourth Edition. *J Pain Symptom Manage*. *56*(6), 831-870.

Anhang Price R, Stucky B, Parast L, Elliott MN, Haas A, Bradley M, Teno JM. (2018). Development of Valid and Reliable Measures of Patient and Family Experiences of Hospice Care for Public Reporting. *J Pall Med.* 21(7):924-932.

Bernacki RE, Block SD, American College of Physicians High Value Care Task Force. (2014). Communication about serious illness care goals: A review and synthesis of best practices. *JAMA Intern Med.*174:1994–2003.

Detering KM, Hancock AD, Reade MC, Silvester W. (2010). The impact of advance care planning on end of life care in elderly patients: Randomised controlled trial. *BMJ.* 340:c1345.

Gomes B, Calanzani N, Curiale V, et al. (2013) Effectiveness and cost-effectiveness of home palliative care services for adults with advanced illness and their caregivers. *Cochrane Database of Systematic Reviews* 6, Cd007760.

Leung JM, Udris EM, Uman J, Au DH. (2012). The effect of end-of-life discussions on perceived quality of care and health status among patients with COPD. *Chest*. 142:128–133.

Mack JW, Weeks JC, Wright AA, et al. (2010). End-of-life discussions, goal attainment, and distress at the end of life: Predictors and outcomes of receipt of care consistent with preferences. *J Clin Oncol* 2010;28:1203–1208.

Singer AE, Goebel JR, Kim YS, et al. (2016). Populations and interventions for palliative and end-of-life care: a systematic review. *J Pall Med*. 19(9): 995-1008.

Wright AA, Zhang B, Ray A, et al. (2018). Associations between end-of-life discussions, patient mental health, medical care near death, and caregiver bereavement adjustment. *JAMA*. 300:1665–1673.

Zhang B, Wright AA, Huskamp HA, et al. (2009). Health care costs in the last week of life: Associations with end-of-life conversations. *Arch Intern Med*. 169:480–488.

[Response Ends]

1b.01. Briefly explain the rationale for this measure.

Explain how the measure will improve the quality of care, and list the benefits or improvements in quality envisioned by use of this measure.

[Response Begins]

Caring for high-need, high-cost seriously ill people is an urgent priority because they account for an increasing proportion of users of the health care system, their care is expensive, and they are more likely than others to be affected by variations in health care quality and safety due to their frequent use of care (Blumenthal, Chernof et al. 2016).

To address the health care needs of seriously ill people, there has been rapid growth of home-based serious illness programs (Long, Abrams et al., 2017; Cohn, Corrigan et al., 2017; NASEM, 2018; Olson et al., 2019). Although measuring, monitoring, and incentivizing high quality care is critical for this vulnerable population, to date, no standardized measures have been developed for programs that provide care to seriously ill people in their homes. Experts have highlighted the need for standardized measures of patient and family care in such programs (NQF, 2020). Quality measures are critical for the determination of value in both current payment models, such as Medicare Advantage and Medicare accountable care organizations (ACOs), and future alternative payment models under consideration to serve seriously ill populations.

Patient- and family-centeredness is particularly important for seriously ill patients, given great variability across patients with regard to both preferences for care intensity and tradeoffs between quality and length of life. Surveys of patients and their family caregivers are the main means of assessing patient- and family-centeredness of care, and survey results can be used to identify areas of patient and family experiences of care that need improvement, monitor quality over time, and be incorporated into value-based models to allow for benchmarking and comparison of programs (Elliott, Cohea, et al., 2015; Rodriguez, von Glahn, et al., 2009; Quigley and McCleskey, 2020).

Expert consensus has identified core areas of high-quality serious illness care about which patients and families are the best and only source of information, including person- and family-centered communication, access and responsiveness, shared decision making and care planning in support of patient goals, care coordination (including medication management), symptom management, and attention to caregiver needs and social determinants of health (Long, Abrams et al., 2017; Cohn, Corrigan et al., 2017). In prior studies, seriously ill patients and/or their family caregivers have reported substantial unmet need in each of these areas, including inadequate communication and reports of not feeling “heard or understood” by their health care providers (Ahluwalia et al., 2022a; Casarett et al., 2008; Covinsky et al., 2000; Dy et al., 2008; Teno et al., 2004), not getting help as soon as needed or on evenings and weekends (Anhang Price et al., 2018), receiving care that is not consistent with their preferences (Khandelwal et al., 2017; Teno et al., 2015), and not receiving desired help with pain and other symptoms (Ahluwalia et al., 2022b; Anhang Price et al., 2018; Teno et al., 2004; Teno et al., 2015).

Citations:

Ahluwalia SC, Vegetabile BG, Edelen MO, Setodji CM, Rodriguez A, et al. (2022a). MACRA Palliative Care Quality Measure Development – Testing Summary Report. Measure Name: Feeling Heard and Understood. *RAND Health Q.* 9(3): 3.

Ahluwalia SC, Vegetabile BG, Edelen MO, Setodji CM, Rodriguez A, et al. (2022b). MACRA Palliative Care Quality Measure Development – Testing Summary Report. Measure Name: Receiving Desired Help for Pain. *RAND Health Q.* 9(3): 3.

Anhang Price R, Stucky B, Parast L, Elliott MN, Haas A, Bradley M, Teno JM. (2018). Development of Valid and Reliable Measures of Patient and Family Experiences of Hospice Care for Public Reporting. *J Pall Med.* 21(7):924-932.

Blumenthal, D., B. Chernof, T. Fulmer, J. Lumpkin and J. Selberg (2016). "Caring for High-Need, High-Cost Patients - An Urgent Priority." *N Engl J Med.* 375(10): 909-911.

Casarett, D., Pickard, A., Bailey, F. A., Ritchie, C. S., Furman, C. D., Rosenfeld, K., Shreve, S., & Shea, J. (2008). A nationwide VA palliative care quality measure: the family assessment of treatment at the end of life. *J Palliat Med.* 11(1), 68-75.

Cohn J, Corrigan J, Lynn J, et al. (2017). Community-Based Models 20

Covinsky, K. E., Fuller, J. D., Yaffe, K., Johnston, C. B., Hamel, M. B., Lynn, J., Teno, J. M., & Phillips, R. S. (2000). Communication and decision-making in seriously ill patients: findings of the SUPPORT project. The Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments. *J Am Geriatr Soc.* 48(5 Suppl), S187-193.

Dy, S. M., Shugarman, L. R., Lorenz, K. A., Mularski, R. A., & Lynn, J. (2008). A systematic review of satisfaction with care at the end of life. *J Am Geriatr Soc.* 56(1), 124-129.

Elliott MN, Cohea CW, Lehrman WG, et al. (2015). Accelerating improvement and narrowing gaps: Trends in patients' experiences with hospital care reflected in HCAHPS public reporting. *Health Serv Res.* 50(6):1850-1867. 10.1111/1475-6773.12305.

Khandelwal, N., J. R. Curtis, V. A. Freedman, J. D. Kasper, P. Gozalo, R. A. Engelberg and J. M. Teno. (2017). "How Often Is End-of-Life Care in the United States Inconsistent with Patients' Goals of Care?" *J Palliat Med.* 20(12): 1400-1404.

Long P, Abrams M, Milstein A, et al., eds. (2017). Effective Care for High-Need Patients: Opportunities for Improving Outcomes, Value, and Health. Washington, DC: National Academy of Medicine.

National Academies of Sciences, Engineering, and Medicine. (2018). Models and Strategies to Integrate Palliative Care Principles into Care for People with Serious Illness: Proceedings of a Workshop. Washington, DC: The National Academies Press.

National Quality Forum (NQF). (2020). Issue Brief: Opportunities for Advancing Quality Measurement in Community-Based Serious Illness Care. Available at: https://store.qualityforum.org/products/issue-brief-opportunities-or-advancing-quality-measurement-in-community-based-serious-illness-care.

Olson A, Harker M, Saunders R, et al. (2019) Innovations in Medicare Advantage To Improve Care for Seriously Ill Patients: Duke Margolis Center for Health Policy, Robert Wood Johnson Foundation.

Quigley DD, McCleskey SG. (2020). Improving care experiences for patients and caregivers at end of life: A systematic review. *Am J Hosp Palliat Care*. 38(1):84-93.

Rodriguez HP, von Glahn T, Elliott MN, et al. (2009). The effect of performance-based financial incentives on improving patient care experiences: A statewide evaluation. *J Gen Intern Med*. 24(12):1281-1288.

Teno, J. M., Clarridge, B. R., Casey, V., Welch, L. C., Wetle, T., Shield, R., & Mor, V. (2004). Family perspectives on end-of-life care at the last place of care. *JAMA*, 291(1), 88-93.

Teno, J. M., Freedman, V. A., Kasper, J. D., Gozalo, P., & Mor, V. (2015). Is care for the dying improving in the United States? *J Palliat Med.* 18(8), 662-666.

[Response Ends]

1b.02. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis.

Include mean, std dev, min, max, interquartile range, and scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

Table 1b.2a displays the mean, standard deviation, 25th, 50th and 75th percentile scores for each of the multi-item and single-item Serious Illness Survey for Home-Based Programs measures. Scores are for the 28 programs participating in the field test that had 10 or more completed surveys.

The measure with the highest top-box score was Communication (top-box mean=78.5; IQR: 9.9). The measure with the lowest top-box scores was Planning for Care (top-box mean=55.5; IQR: 8.6). Across all measures, there is substantial room for improvement.

Table 1b.02a. Distribution of Top-Box Scores for Serious Illness Survey for Home-Based Programs Measures

*	Mean	Standard Deviation	25th Percentile	50th Percentile	75th Percentile
Multi-item Measures	*	*	*	*	*
Communication	78.5	7.8	73.1	78.8	83.0
Care Coordination	68.1	8.2	62.8	67.1	75.6
Help for Symptoms	60.0	10.8	51.2	59.2	65.5
Planning for Care	55.5	9.4	50.5	54.6	59.0
Support for Family and Friends	70.8	10.9	64.0	71.0	78.4
Single-item Measures	*	*	*	*	*
Overall Rating of Program	75.3	9.2	69.1	73.2	78.8
Willingness to Recommend Program	73.7	10.7	66.4	74.8	80.0

Table 1b.02a displays the distribution of top-box scores for the seven Serious Illness Survey for Home-Based Programs quality measures, including the mean, standard deviation, 25th, 50th and 75th percentiles of program performance in the survey field test.

* Cell intentionally left blank.

Additional detail regarding current performance of the measures, including minimum, maximum, and decile scores, is attached in the Supplementary Appendix, “Table 1b.02b. Serious Illness Survey for Home-Based Programs Measures Deciles 2022_11."

[Response Ends]

1b.03. If no or limited performance data on the measure as specified is reported above, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement. Include citations.

[Response Begins]

N/A

[Response Ends]

1b.04. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability.

Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included. Include mean, std dev, min, max, interquartile range, and scores by decile. For measures that show high levels of performance, i.e., “topped out”, disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b) under Usability and Use.

[Response Begins]

Using data from 2,263 respondents to the 2019 field test of the Serious Illness Survey for Home-Based Programs, we find that male patients report care experiences that are statistically significantly different than females on one measure (Willingness to Recommend the Program), reporting that they are 4 points less likely to "definitely" recommend the program. Black patients report worse care experiences than White patients for six of seven measures, with differences ranging from one to nine points across these measures (reaching statistical significance for three measures), and Hispanic patients report worse care experiences than White patients on all seven measures with differences between four and 11 points across measures (reaching statistical significance for five measures). Younger patients report better care experiences than those age 90 and older; for patients 18-54, these differences range from five to 15 points across measures and are statistically significant for five of seven measures. Patients with Medicaid report statistically significantly better experiences than those with Medicare for one measure (Willingness to Recommend; 10-point difference), with no other significant differences between these groups across measures.

[Response Ends]

1b.05. If no or limited data on disparities from the measure as specified is reported above, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in above.

[Response Begins]

N/A

[Response Ends]

2. Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.

---

sp.01. Provide the measure title.

Measure titles should be concise yet convey who and what is being measured (see What Good Looks Like).

[Response Begins]

Serious Illness Survey for Home-Based Programs

[Response Ends]

sp.02. Provide a brief description of the measure.

Including type of score, measure focus, target population, timeframe, (e.g., Percentage of adult patients aged 18-75 years receiving one or more HbA1c tests per year).

[Response Begins]

The proposed measures are derived from the Serious Illness Survey for Home-Based Programs, a 36-item questionnaire designed to measure the care experiences of patients receiving care from home-based serious illness programs. Home-based serious illness programs provide care for seriously ill patients at their private residences (i.e., in their homes or assisted living facilities, not in institutions like skilled nursing facilities). Programs are staffed by interdisciplinary teams that provide support for palliation of symptoms, assist with coordination of care, answer questions after-hours, provide medication management, and assist with advance care planning (Cohn et al., 2017). Teams consist of clinicians (e.g. physicians, nurse practitioners) that oversee care, as well as clinical and supportive staff that make home visits (e.g. registered nurses, social workers, CNAs). Programs serve patients with a life expectancy that ranges from 1-5 years and have enrollment criteria based on diagnosis, symptom burden, functional status, and/or prior health care utilization.

The five proposed multi-item measures are:

1. Communication
2. Care Coordination
3. Help for Symptoms
4. Planning for Care
5. Support for Family and Friends

The two proposed single-item measures are:

1. Overall Rating of the Program
2. Willingness to Recommend the Program

Appendix A presents the survey items included in each measure, including response options for each item. Measure scores are “top-box” scores that reflect the percent of respondents who select the most positive response category(ies) in response to the survey item(s) within the measure.

Citation:

Cohn J, Corrigan J, Lynn J, Meier D, et al. Community-Based Models of Care Delivery for People with Serious Illness. National Academy of Medicine Discussion Paper. Available at https://nam.edu/wp-content/uploads/2017/04/Community-Based-Models-of-Care-Delivery-for-People-with-Serious-Illness.pdf.

[Response Ends]

sp.04. Check all the clinical condition/topic areas that apply to your measure, below.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Surgery: General*

[Response Begins]

Palliative Care and End-of-Life Care

[Response Ends]

sp.05. Check all the non-condition specific measure domain areas that apply to your measure, below.

[Response Begins]

Person-and Family-Centered Care: Person-and Family-Centered Care

[Response Ends]

sp.06. Select one or more target population categories.

Select only those target populations which can be stratified in the reporting of the measure's result.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Populations at Risk: Populations at Risk*

[Response Begins]

Elderly (Age >= 65)

Women

[Response Ends]

sp.07. Select the levels of analysis that apply to your measure.

Check ONLY the levels of analysis for which the measure is SPECIFIED and TESTED.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*
- *Population: Population*

[Response Begins]

Other

[Response Ends]

sp.08. Indicate the care settings that apply to your measure.

Check ONLY the settings for which the measure is SPECIFIED and TESTED.

[Response Begins]

Home Care

[Response Ends]

sp.09. Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials.

Do not enter a URL linking to a home page or to general information. If no URL is available, indicate "none available".

[Response Begins]

https://www.rand.org/health-care/surveys_tools/serious-illness-survey.html

[Response Ends]

sp.12. Attach the data dictionary, code table, or value sets (and risk model codes and coefficients when applicable). Excel formats (.xlsx or .csv) are preferred.

Attach an excel or csv file; if this poses an issue, contact staff. Provide descriptors for any codes. Use one file with multiple worksheets, if needed.

[Response Begins]

No data dictionary/code table – all information provided in the submission form

[Response Ends]

For the question below: state the outcome being measured. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.13. State the numerator.

Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome).

DO NOT include the rationale for the measure.

[Response Begins]

Measure scores are "top-box" scores that reflect the percent of respondents who select the most positive response category(ies) in response to the survey item(s) within the measure. Therefore, the numerator is the number of respondents who select the most positive response category(ies) in response to the survey items within the measure.

[Response Ends]

For the question below: describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.14. Provide details needed to calculate the numerator.

All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

For each survey item, the numerator is the number of respondents who selected the most positive response category(ies), as described in Table 1a.

Table 1a. Top-Box Response Categories for Each Response Scale on the Serious Illness Survey for Home-Based Programs

Response Scale	Top-Box Response (most positive)
"Always," "Usually," "Sometimes," "Never"	Always
"Yes, definitely," "Yes, somewhat," "No"	Yes, definitely
"Definitely yes," "Probably yes," "Probably no," "Definitely no"	Definitely yes
Rating 0–10, where 10 is the most positive	9 or 10

Table shows the response categories that are the most positive for each response scale, and therefore, are used as the numerator when calculating top-box scores.

To calculate the numerator, first calculate the "top-box" score for each item within the measure by assigning a score of "100" if the most-positive response category(ies) for that item is selected, and a score of "0" if any other response category is selected. Then, adjust each item score for mode of administration (e.g., mail-only or mail-telephone) for each respondent. Then, calculate the case-mix adjusted score for each item for each program. Then, take the unweighted means of the mode- and case-mix-adjusted program-level items to calculate the multi-item measure score.

[Response Ends]

For the question below: state the target population for the outcome. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.15. State the denominator.

Brief, narrative description of the target population being measured.

[Response Begins]

Survey respondents are patients receiving care from home-based serious illness programs. Survey eligibility criteria and exclusions are detailed below in sections sp.16 – sp.18. Screener questions and tailored non-applicable

response options (e.g., I did not want help for my pain) are used to identify respondents who are and are not eligible to respond to survey items included in evaluative measures. Therefore, denominators vary by survey item (and corresponding multi-item measures, if applicable) according to the eligibility of respondents for each item.

[Response Ends]

For the question below: describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in sp.22.

sp.16. Provide details needed to calculate the denominator.

All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets.

Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

The survey population is patients receiving care from a home-based serious illness program. For each item in a measure, the denominator is the number of respondents who answered the item. Therefore, the denominator for single-item measures is the number of respondents who answered the item. The denominator for multi-item measures is the number of respondents who answer at least one item within the multi-item measure. Multi-item measure scores are the average proportion of respondents that gave responses in the most positive category across the items in the multi-item measure (as discussed above in sp.14).

[Response Ends]

sp.17. Describe the denominator exclusions.

Brief narrative description of exclusions from the target population.

[Response Begins]

The Serious Illness Survey for Home-Based Programs is designed for administration to adult patients who are currently enrolled in home-based serious illness programs. Patients are excluded from the survey sample if they:

- Are under age 18
- Receive care from a serious illness program in a setting OTHER than home or an assisted living facility (e.g., in a nursing home or other long-term care facility)
- Are known to have been discharged to hospice
- Are known to have died
- Have been enrolled in the serious illness program for less than six weeks as of the date of survey sampling

In keeping with the Medicare CAHPS Survey (https://www.cms.gov/files/document/ma-pdp-cahps-qapts-v11-complete-manual.pdf), a survey is considered partially completed if there are responses to at least one measure and for less than 50 percent of survey items that are applicable to all. A survey is considered completed if there are responses to at least one measure and for 50 percent or more of the survey items that are applicable to all. Final analytic datasets include all completed and partially completed surveys.

There are no explicit exclusions based on language; the survey is available in English and Spanish.

[Response Ends]

sp.18. Provide details needed to calculate the denominator exclusions.

All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at sp.11.

[Response Begins]

Serious illness programs can use data from their administrative records to identify denominator exclusions prior to sampling. All programs participating in the field test were readily able to provide data to identify patients meeting inclusion/exclusion criteria.

No survey questions are used to further screen patients for survey eligibility.

If it is learned during mail or telephone data collection that a patient died between the date the sample was drawn and the date the survey was administered, the case should be marked as ineligible and no further contact should be made.

[Response Ends]

sp.19. Provide all information required to stratify the measure results, if necessary.

Include the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate. Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format in the Data Dictionary field.

[Response Begins]

N/A. Measures are designed to calculate program-level scores, not for stratification.

[Response Ends]

sp.20. Is this measure adjusted for socioeconomic status (SES)?

[Response Begins]

Yes

[Response Ends]

sp.21. Select the risk adjustment type.

Select type. Provide specifications for risk stratification and/or risk models in the Scientific Acceptability section.

[Response Begins]

Statistical risk model

[Response Ends]

sp.22. Select the most relevant type of score.

Attachment: If available, please provide a sample report.

[Response Begins]

Rate/proportion

[Response Ends]

sp.23. Select the appropriate interpretation of the measure score.

Classifies interpretation of score according to whether better quality or resource use is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score

[Response Begins]

Better quality = Higher score

[Response Ends]

sp.24. Diagram or describe the calculation of the measure score as an ordered sequence of steps.

Identify the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period of data, aggregating data; risk adjustment; etc.

[Response Begins]

To calculate top-box scores for Serious Illness Survey for Home-Based Programs measures:

1. Identify the target respondent population (as described above in sp.16).
2. Identify any exclusions from the respondent population (as described above in sp.17).
3. Score each item using top-box method (possible values of 0 or 100, as described above in sp.14).
4. Calculate mode-adjusted top-box scores for each item.
5. Calculate case-mix adjusted top-box scores for each item for each program, using a linear regression-based approach that adjusts for all variables listed in 2b.20.
6. Average top-box scores across the items within each multi-item measure, weighting each item equally. If data are missing for a respondent for an item(s) within a multi-item measure, the respondent's answers to other items within the measure are still used in the calculation of multi-item measure scores.

[Response Ends]

sp.25. Attach a copy of the instrument (e.g. survey, tool, questionnaire, scale) used as a data source for your measure, if available.

[Response Begins]

Copy of instrument is attached.

[Response Ends]

sp.26. Indicate the responder for your instrument.

[Response Begins]

Patient

[Response Ends]

sp.27. If measure testing is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.

Examples of samples used for testing:

• Testing may be conducted on a sample of the accountable entities (e.g., hospital, physician). The analytic unit specified for the particular measure (e.g., physician, hospital, home health agency) determines the sampling strategy for scientific acceptability testing.

• The sample should represent the variety of entities whose performance will be measured. The 2010 Measure Testing Task Force recognized that the samples used for reliability and validity testing often have limited generalizability because measured entities volunteer to participate. Ideally, however, all types of entities whose performance will be measured should be included in reliability and validity testing.

• The sample should include adequate numbers of units of measurement and adequate numbers of patients to answer the specific reliability or validity question with the chosen statistical method.

• When possible, units of measurement and patients within units should be randomly selected.

[Response Begins]

The Serious Illness Survey for Home-Based Programs should be administered to either (1) all patients of a given serious illness program (a census) or (2) a simple random sample of patients from the program. Should users wish to make comparisons across programs, we recommend that users set a minimum sample size such that at least 100 completed surveys per program will be achieved. This will allow for sufficient precision to distinguish between programs' performance. However, even if this threshold is not met, survey responses are useful for understanding patient experiences within the program and for tracking changes within the program over time.

Please note: the minimum sample size of 100 surveys per program is to ensure program level reliability, e.g., the ability to reliably distinguish between serious illness programs' performance, as described in Section 2a.11. However, when testing measures and conducting psychometric analyses, sample size requirements are related to power and precision and to limit bias. For testing, we impose a minimum of 10 per program to be as inclusive of reporting units as possible while avoiding bias in estimating variance components that comes from including units with very small sample sizes.

[Response Ends]

sp.28. Identify whether and how proxy responses are allowed.

[Response Begins]

To allow for reports of care experiences for patients whose functional status or cognitive ability may make it difficult to complete the survey on their own, proxy responses are permitted. The cover of the mail survey instrument and introduction to the telephone instrument provide instructions for proxy respondents, noting that survey questions refer to the patient's care and that proxy respondents should respond regarding the patient's experience rather than their own experience, unless a question indicates otherwise. Proxy respondents should be individuals who are knowledgeable about the care the patient received from the program. Employees of the program should not be proxy respondents or help patients complete the survey. Only individuals within the sampled patient's household are included as telephone proxy respondents. The survey instrument collects information about whether and how a proxy respondent assisted in completing the survey; proxy assistance is one of the recommended variables for case-mix adjustment of quality measure scores, as described in 2b.20.

[Response Ends]

sp.29. Survey/Patient-reported data.

Provide instructions for data collection and guidance on minimum response rate. Specify calculation of response rates to be reported with performance measure results.

[Response Begins]

The survey was field-tested in two modes: (1) a mail-only mode, which consisted of a prenotification letter, followed by a mail survey one week later, and an additional mail survey three weeks after that if the survey had not been returned, and (2) a mail-telephone mode, which consisted of a prenotification letter, followed by a mail survey one week later, and up to five calls to complete the survey by phone if the mail survey was not returned after three weeks. Consistent with the findings for other patient experience surveys, the mail-telephone mode of survey administration yielded a much higher response rate (42.5%) than mail only (30.4%), and mixed mode was especially effective in increasing response rates among patients with Medicaid as the primary payer, more than doubling their response rates (DeYoreo et al., 2022). These findings suggest that when feasible, use of mail-telephone administration can help programs to achieve higher response rates and broader representation. Detailed information regarding survey administration in each of the two field-tested modes is available in the Guidelines for Survey Administration and Analysis on the survey website at: https://www.rand.org/health-care/surveys_tools/serious-illness-survey.html.

Response rates should be calculated using the mode-appropriate American Association for Public Opinion Research response rate calculator (https://www.aapor.org/Education-Resources/For-Researchers/Poll-Survey-FAQ/Response-Rates-An-Overview.aspx). There is no recommended minimum response rate; rather, users are encouraged to choose an affordable mode of survey administration that maximizes survey response rate.

The survey is available in English and Spanish.

Citation:

DeYoreo M, Anhang Price R, Bradley MA, Schlang D, Montemayor CK, Tolpadi A, Cleary PD, Teno JM, and Elliott MN. Adding telephone follow-up can improve representativeness of surveys of seriously ill people. *J Am Geratr Soc.* 2022 Jun;70(6):1870-1873.

[Response Ends]

sp.30. Select only the data sources for which the measure is specified.

[Response Begins]

Instrument-Based Data

[Response Ends]

sp.31. Identify the specific data source or data collection instrument.

For example, provide the name of the database, clinical registry, collection instrument, etc., and describe how data are collected.

[Response Begins]

Serious Illness Survey for Home-Based Programs

[Response Ends]

sp.32. Provide the data collection instrument.

[Response Begins]

Available in attached appendix in Question 1 of the Additional Section

[Response Ends]

Measure testing must demonstrate adequate reliability and validity in order to be recommended for endorsement. Testing may be conducted for data elements and/or the computed measure score. Testing information and results should be entered in the appropriate fields in the Scientific Acceptability sections of the Measure Submission Form.

- Measures must be tested for all the data sources and levels of analyses that are specified. If there is more than one set of data specifications or more than one level of analysis, contact NQF staff about how to present all the testing information in one form.
- All required sections must be completed.
- For composites with outcome and resource use measures, Questions 2b.23-2b.37 (Risk Adjustment) also must be completed.
- If specified for multiple data sources/sets of specifications (e.g., claims and EHRs), Questions 2b.11-2b.13 also must be completed.
- An appendix for supplemental materials may be submitted (see Question 1 in the Additional section), but there is no guarantee it will be reviewed.
- Contact NQF staff with any questions. Check for resources at the Submitting Standards webpage.
- For information on the most updated guidance on how to address social risk factors variables and testing in this form refer to the release notes for the 2021 Measure Evaluation Criteria and Guidance.

Note: The information provided in this form is intended to aid the Standing Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF's evaluation criteria for testing.

2a. Reliability testing demonstrates the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise. For instrument-based measures (including PRO-PMs) and composite performance measures, reliability should be demonstrated for the computed performance score.

2b1. Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For instrument based measures (including PRO-PMs) and composite performance measures, validity should be demonstrated for the computed performance score.

2b2. Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure;

AND

If patient preference (e.g., informed decision-making) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).

2b3. For outcome measures and other measures when indicated (e.g., resource use):

- an evidence-based risk-adjustment strategy (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at start of care; 14,15 and has demonstrated adequate discrimination and calibration
- rationale/data support no risk adjustment/ stratification.

2b4. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful 16 differences in performance;

OR

there is evidence of overall less-than-optimal performance.

2b5. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b6. Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders) and how the specified handling of missing data minimizes bias.

2c. For composite performance measures, empirical analyses support the composite construction approach and demonstrate that:

2c1. the component measures fit the quality construct and add value to the overall composite while achieving the related objective of parsimony to the extent possible; and

2c2. the aggregation and weighting rules are consistent with the quality construct and rationale while achieving the related objective of simplicity to the extent possible.

(if not conducted or results not adequate, justification must be submitted and accepted)

Definitions

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measures scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality. The degree of consensus and any areas of disagreement must be provided/discussed.

Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

Risk factors that influence outcomes should not be specified as exclusions.

With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of $25 in cost for an episode of care (e.g., $5,000 v.$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

Please separate added or updated information from the most recent measure evaluation within each question response in the Scientific Acceptability sections. For example:

Current Submission:

Updated testing information here.

Previous (Year) Submission:

Testing from the previous submission here.

2a.01. Select only the data sources for which the measure is tested.

[Response Begins]

Instrument-Based Data

[Response Ends]

2a.02. If an existing dataset was used, identify the specific dataset.

The dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry).

[Response Begins]

Survey data are from a field test of the Serious Illness Survey for Home-Based Programs conducted from October 2019 through January 2020 among patients of 32 serious illness programs. Patients within each program were eligible for the survey if they were adults (age 18 or older at the time the sample was selected), received care at a private home or assisted living facility, and had been receiving care from the program for at least three months and no more than 24 months at the time the sample was selected. Sampled patients who indicated on the survey that they had not received visits from the program in the past three months were not considered eligible.

Within each program, eligible patients were randomly assigned to one of two modes of survey administration, mail-only or mail-telephone. The mail-only mode consisted of a pre-notification letter, followed by a mail survey one week later, and an additional mail survey three weeks after that if the survey had not been returned. The mail-telephone mode consisted of a pre-notification letter, followed by a mail survey one week later, and up to five calls to complete the survey by phone if the mail survey was not returned after three weeks. The survey was available in both English and Spanish; when indicated in the sample file, Spanish was the language for mailing and initial telephone calls. Spanish was also offered as an option by telephone interviewers.

A total of 2,263 respondents completed the field test survey and are included in the dataset for the analyses that follow.

[Response Ends]

2a.03. Provide the dates of the data used in testing.

Use the following format: "MM-DD-YYYY - MM-DD-YYYY"

[Response Begins]

10-24-2019 – 01-07-2020

[Response Ends]

2a.04. Select the levels of analysis for which the measure is tested.

Testing must be provided for all the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan.

Please refrain from selecting the following answer option(s). We are in the process of phasing out these answer options and request that you instead select one of the other answer options as they apply to your measure.

Please do not select:

- *Clinician: Clinician*

- *Population: Population*

[Response Begins]

Other (specify)

[Other (specify) Please Explain]

Home-based serious illness program. The level of analysis is "serious illness program" for all analyses shown. As noted in 2a).05, serious illness programs may be owned and operated by a variety of entity types, such as hospices, health systems, or medical groups. However, regardless of ownership, serious illness programs provide similar types of services, use the same types of staff, and serve similar populations. Just as a hospital might be owned by a health system, but is its own accountable entity, so too, a serious illness program is an entity for quality assessment.

[Response Ends]

2a.05. List the measured entities included in the testing and analysis (by level of analysis and data source).

Identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample.

[Response Begins]

The study was conducted in 32 geographically diverse serious illness programs that provide home-based care across the United States. Programs were recruited from a master list of 319 serious illness programs developed by the project team, primarily from the Center to Advance Palliative Care (CAPC) National Palliative Care Registry (CAPC 2021), a report on community-based model programs for the seriously ill (Kerr 2017), and responses to announcements posted by the National Hospice & Palliative Care Organizationand the American Academy of Hospice and Palliative Medicine. To be eligible to participate, all programs needed to provide medical care to seriously ill patients in their homes. Almost all included programs provide after-hours access to care either by phone or in person and have either a physician or a nurse practitioner on the team that makes home visits. As described in Table 2a.1, approximately one-fifth of participating programs operate in the Northeast, South, and regions, respectively, 28% operate in the West, and nine percent operate in more than one region. One quarter of programs had 250 or more patients actively in care at a given time, while 28% had fewer than 50 in care. Seventy-two percent of the programs are owned and operated out of hospices or home health agencies, 6 out of health systems or health plans, and 3 are part of medical groups. (Please note that while serious illness programs may be owned and operated by a variety of entity types, such as hospices, health systems, or medical groups, regardless of ownership, serious illness programs provide similar types of services, use the same types of staff, and serve similar populations.)

Table 2a.1. Characteristics of Programs Participating in the Serious Illness Survey for Home-Based Programs Field Test, N=32

Characteristic	n (%)
Census region: Northeast	6 (18.8%)
Census region: South	7 (21.9%)
Census region: Midwest	7 (21.9%)
Census region: West	9 (28.1%)
Census region: More than one region	3 (9.4%)
Number of patients in care: <50	9 (28.1%)

Characteristic	n (%)
Number of patients in care: 50-99	6 (18.8%)
Number of patients in care: 100-<250	9 (28.1%)
Number of patients in care: 250+	8 (25.0%)
Ownership: Hospice or home health-based	23 (71.9%)
Ownership: Health system-based	6 (18.8%)
Ownership: Medical group-based	3 (9.4%)

Table shows characteristics of the 32 home-based serious illness programs participating in the Serious Illness Survey for Home-Based Programs field test, including census region, number of patients in care, and ownership.

Citations:

Center to Advance Palliative Care: National Palliative Care Registry™. Available at: https://www.capc.org/national-palliative-care-registry/ Accessed July 20, 2022.

Kerr K: Community-Based Model Programs for the Seriously Ill. The Gordon and Betty Moore Foundation. May 2017. Available at: https://www.moore.org/docs/default-source/patient-care-/report-model-programs-for-the-seriously-ill-may-2017-dls.%20pdf?sfvrsn=529b6c0c_2 Accessed July 20, 2022.

[Response Ends]

2a.06. Identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis), separated by level of analysis and data source; if a sample was used, describe how patients were selected for inclusion in the sample.

If there is a minimum case count used for testing, that minimum must be reflected in the specifications.

[Response Begins]

Patients within each program were eligible for the survey if they were adults (age 18 or older at the time the sample was selected), received care at a private home or assisted living facility, and had been receiving care from the program for at least three months and no more than 24 months at the time the sample was selected. Patients who indicated on the survey that they had not received visits from the program in the past three months were not considered eligible. We took a census of patients meeting these eligibility criteria from each program participating in the field test.

As shown in Table 2a.2, 58 percent of respondents were female. Fourteen percent of respondents were under age 65; nearly 40 percent were age 85 or older. Three quarters of respondents were non-Hispanic white, 10 percent were Black or African American, and 10 percent were Hispanic. Forty-four percent received seven or more in-person visits from the program in three months. Fourteen percent reported that they were not able to leave the house, and 58% reported being in fair or poor health. Proxy respondents completed the survey on behalf of the patient for 33% of respondents; an additional 14% reported some other form of proxy assistance.

Table 2a.2. Characteristics of Patients Responding to Serious Illness Survey for Home-Based Programs Field Test, N=2,263.

Characteristic	Respondents n (%)[a] n=2,263
Sex: Female	1,322 (58.4%)
Sex: Male	941 (41.6%)

Characteristic	Respondents n (%)[a] n=2,263
Age: 18-64	312 (13.8%)
Age: 65-74	401 (17.7%)
Age: 75-84	670 (29.6%)
Age: 85-89	441 (19.5%)
Age: 90 or older	439 (19.4%)
Race/ Ethnicity: Black or African American	226 (10.0%)
Race / Ethnicity: Hispanic	235 (10.4%)
Race / Ethnicity: Other	95 (4.2%)
Race / Ethnicity: White	1,706 (75.4%)
Length of stay in program (through beginning of survey administration): 6 weeks to 3 months	300 (13.3%)
Length of stay in program: 4 to less than 6 months	436 (19.3%)
Length of stay in program: 6 to less than 12 months	769 (34.0%)
Length of stay in program: 12 to 24 months	758 (33.5%)
Primary diagnosis: Cancer	307 (13.6%)
Primary diagnosis: Alzheimer's or dementia	212 (9.4%)
Primary diagnosis: All other	1,745 (77.1%)
Number of In-Person Visits in 3 Months: 1 to 2 times	287 (12.7%)
Number of In-Person Visits in 3 Months: 3 to 4 times	25 (11.4%)
Number of In-Person Visits in 3 Months: 5 to 6 times	235 (10.4%)
Number of In-Person Visits in 3 Months: 7 or more times	993 (43.9%)
Number of In-Person Visits in 3 Months: Missing	491 (21.7%)
Proxy assistance with survey response: Proxy completed survey for patient	753 (33.3%)
Proxy assistance with survey response: Proxy assisted in some other way	316 (14.0%)
Proxy assistance with survey response: No proxy assistance	1,194 (52.8%)
Self-Reported Functional Status: Able to leave house	1,944 (85.9%)
Self-Reported Functional Status: Able to get out of bed but not house	167 (7.4%)
Self-Reported Functional Status: Not able to get out of bed	152 (6.7%)
Self-Reported Physical Health Status: Excellent	52 (2.3%)
Self-Reported Physical Health Status: Very good	226 (10.0%)
Self-Reported Physical Health Status: Good	668 (29.5%)
Self-Reported Physical Health Status: Fair	883 (39.0%)
Self-Reported Physical Health Status: Poor	434 (19.2%)
Self-Reported Mental Health Status: Excellent	253 (11.2%)

Characteristic	Respondents n (%)[a] n=2,263
Self-Reported Mental Health Status: Very good	507 (22.4%)
Self-Reported Mental Health Status: Good	756 (33.4%)
Self-Reported Mental Health Status: Fair	550 (24.3%)
Self-Reported Mental Health Status: Poor	197 (8.7%)

Table shows characteristics of the 2,263 patients responding to the Serious Illness Survey for Home-Based Programs field test, including sex, age, race/ethnicity, length of stay, primary diagnosis, number of visits in the last 3 months, proxy assistance with the survey, and self-reported functional and health status.

[a] Percentages were calculated among non-missing values, except where a large missing category is noted (i.e., number of in-person visits).

[Response Ends]

2a.07. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing.

[Response Begins]

N/A. All testing was conducted using the dataset described in 2a.02.

[Response Ends]

2a.08. List the social risk factors that were available and analyzed.

For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

[Response Begins]

Social risk factors in the data include patient education, language spoken at home, and primary payer for care in the serious illness program. As shown in Table 2a.3., 21 percent of survey respondents had no high school degree and 37 percent had a high school degree or GED; seven percent spoke Spanish or another language other than English at home. Seven percent of respondents had Medicaid as their primary payer.

Table 2a.3. Social Risk Factor Characteristics of Patients Responding to Serious Illness Survey for Home-Based Programs Field Test, N=2,263.

Characteristic	Respondents n (%)[a] n=2,263
Education: No high school degree	480 (21.2%)
Education: High school graduate or GED	835 (36.9%)
Education: Some college or 2-year degree	527 (23.3%)
Education: College degree or more	421 (18.6%)
Language spoken at home: English	2,098 (92.7%)
Language spoken at home: Spanish	95 (4.2%)
Language spoken at home: Other	70 (3.1%)

Characteristic	Respondents n (%)[a] n=2,263
Primary payer for care: Medicare (including FFS and Medicare Advantage)	1,084 (47.9%)
Primary payer for care: Medicaid	164 (7.2%)
Primary payer for care: Private	530 (23.4%)
Primary payer for care: Other (including uninsured and no payer) or missing	484 (21.4%)

Social risk factor characteristics of the 2,263 patients responding to the Serious Illness Survey for Home-Based Programs field test, including education, language spoken at home, and primary payer for care.

[a] Percentages were calculated among non-missing values, except where a large missing category is noted (i.e., primary payer for care).

[Response Ends]

Note: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a.09 check patient or encounter-level data; in 2a.010 enter "see validity testing section of data elements"; and enter "N/A" for 2a.11 and 2a.12.

2a.09. Select the level of reliability testing conducted.

Choose one or both levels.

[Response Begins]

Patient or Encounter-Level (e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements)

Accountable Entity Level (e.g., signal-to-noise analysis)

[Response Ends]

2a.10. For each level of reliability testing checked above, describe the method of reliability testing and what it tests.

Describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used.

[Response Begins]

To assess reliability of the proposed quality measures, we calculated the internal consistency reliability of multi-item measures using Cronbach's alpha. Cronbach's alpha increases with the number of items included in a composite measure and their average correlation with one another. Larger values indicate more precise measurement of the underlying construct. Cronbach's alphas of 0.70 or higher are considered adequate for group comparisons (Nunnally and Bernstein, 1994). We also calculated the Pearson item-total correlation, the correlation between a given item and the total multi-item measure score with the given item removed. This metric reflects how related each item is to all other items in the measure.

We also calculated the inter-unit (i.e., program-level) reliability of each measure, a 0-1 index of the degree to which measure scores are able to precisely distinguish between the performances of programs. We calculated the program-level reliability for each measure using intra-class correlations (ICCs) of the case-mix- and survey mode-

adjusted top-box scores, restricting the analysis to programs with 10 or more respondents (28 out of 32 programs). We also calculated predicted program-level reliability with 100 respondents using the Spearman Brown formula (Allen and Yen, 1979). When programs are being compared, measure reliability of 0.70 or greater is commonly considered adequate (Nunnally and Bernstein, 1994). (As noted above, the unit of interest is the home-based serious illness program. Although these programs may be owned and operated by a variety of different entities, regardless of ownership, serious illness programs provide similar types of services, use the same types of staff, and serve similar populations. Just as a hospital might be owned by a health system, but is its own accountable entity, so too, a serious illness program is an entity for quality assessment.)

Citations:

Allen MJ, Yen WM: Introduction to Measurement Theory. Monterey, CA: Brooks/Cole, 1979.

Nunnally JC, Bernstein IH: Psychometric Theory. New York, NY: McGraw-Hill, 1994.

[Response Ends]

2a.11. For each level of reliability testing checked above, what were the statistical results from reliability testing?

For example, provide the percent agreement and kappa for the critical data elements, or distribution of reliability statistics from a signal-to-noise analysis. For score-level reliability testing, when using a signal-to-noise analysis, more than just one overall statistic should be reported (i.e., to demonstrate variation in reliability across providers). If a particular method yields only one statistic, this should be explained. In addition, reporting of results stratified by sample size is preferred (pg. 18, NQF Measure Evaluation Criteria).

[Response Begins]

Table 2a.4. Psychometric Properties of Proposed Serious Illness Survey for Home-Based Programs Quality Measures and Component Items.

Measures & Component Survey Items	Item-total Pearson Correlation	Cronbach's Alpha with Item Deleted (95% CI)
Communication (Cronbach's alpha = 0.85, 95% CI: 0.84, 0.86)	-	-
In the last 3 months, how often did people from this program spend enough time with you when they visited?	0.61	0.83 (0.82, 0.84)
In the last 3 months, how often did people from this program explain things to you in a way you could understand?	0.64	0.82 (0.81, 0.83)
In the last 3 months, how often did people from this program listen carefully to you?	0.69	0.81 (0.79, 0.82)
In the last 3 months, how often did you feel that people from this program cared about you as a whole person?	0.68	0.81 (0.80, 0.83)
In the last 3 months, how often did you feel heard and understood by people from this program?	0.66	0.82 (0.80, 0.83)
Care Coordination (Cronbach's alpha = 0.74, 95% CI: 0.71, 0.77)	-	-

Measures & Component Survey Items	Item-total Pearson Correlation	Cronbach's Alpha with Item Deleted (95% CI)
In the last 3 months, how often did people from this program seem to know the important information about your medical history?	0.51	0.69 (0.66, 0.72)
In the last 3 months, did someone from this program talk with you about the care or treatment you get from your other doctors or health care providers?	0.53	0.69 (0.65, 0.72)
In the last 3 months, did someone from this program talk with you about all the medicines you are taking?	0.54	0.68 (0.64, 0.71)
Everyday activities include things like getting ready in the morning, getting meals, or going places in your community. In the last 3 months, did someone from this program talk with you about how to get help with everyday activities?	0.44	0.72 (0.68, 0.75)
In the last 3 months, when you contacted this program between visits, did you get the help you needed?	0.50	0.70 (0.66, 0.73)
Help for Symptoms (Cronbach's alpha = 0.69, 95% CI: 0.64, 0.74)	-	-
In the last 3 months, did you get as much help as you wanted for your pain?	0.53	0.57 (0.49, 0.65)
In the last 3 months, did you get as much help as you wanted for your breathing?	0.48	0.64 (0.56, 0.71)
In the last 3 months, did you get as much help as you wanted for your feelings of anxiety or sadness?	0.52	0.59 (0.50, 0.66)
Planning for Care (Cronbach's alpha = 0.73, 95% CI: 0.71, 0.74)	-	-
Did someone from this program ever talk with you about what you should do during a health emergency?	0.53	0.67 (0.64, 0.69)
Did someone from this program ever talk with you about what is important in your life?	0.56	0.63 (0.60, 0.65)
Did someone from this program ever talk with you about what your health care options would be if you got sicker?	0.56	0.62 (0.60, 0.65)
Support for Family and Friends (Cronbach's alpha = 0.70, 95% CI: 0.67, 0.73)	-	-
In the last 3 months, did the people from the program involve your family members or friends in discussions about your health care as much as you wanted?	0.54	N/A
In the last 3 months, did your family members or friends get as much emotional support as they wanted from this program?	0.54	N/A
Global Measure: Overall Rating of the Program	-	-
Using any number from 0 to 10, where 0 is the worst care possible and 10 is the best care possible, what number would you use to rate your care from this program?	N/A	N/A

Measures & Component Survey Items	Item-total Pearson Correlation	Cronbach's Alpha with Item Deleted (95% CI)
Global Measure: Willingness to Recommend the Program	-	-
Would you recommend this program to your friends and family?	N/A	N/A

Table showing item-total Pearson correlations between survey items and their measures, and Cronbach's alpha with item deleted for each Serious Illness Survey for Home-Based Programs measure.

Note: CI= Confidence Interval.

- This cell intentionally left empty.

Table 2a.5. Intraclass Correlation and Reliability of Proposed Serious Illness Survey for Home-Based Programs Quality Measures.

Measure	Intraclass Correlation Coefficient (95% CI)	Reliability at 100 Measure Respondents (95% CI)
Multi-item Measures	-	-
Communication (5 items)	0.029 (0.013, 0.076)	0.75 (0.57, 0.89)
Care Coordination (5 items)	0.034 (0.021, 0.097)	0.78 (0.68, 0.91)
Help for Symptoms (3 items)	0.027 (0.012, 0.073)	0.73 (0.54, 0.89)
Planning for Care (3 items)	0.036 (0.018, 0.082)	0.79 (0.64, 0.90)
Support for Family and Friends (2 items)	0.039 (0.018, 0.100)	0.80 (0.64, 0.92)
Global Measures	-	-
Rating of Program (1 item)	0.020 (0.006, 0.058)	0.67 (0.38, 0.86)
Willingness to Recommend (1 item)	0.032 (0.014, 0.076)	0.77 (0.59, 0.89)

Table shows intraclass correlation coefficient and reliability at 100 measure respondents for each Serious Illness Survey for Home-Based Programs measure.

Note: CI= Confidence Interval.

- This cell intentionally left empty.

[Response Ends]

2a.12. Interpret the results, in terms of how they demonstrate reliability.

(In other words, what do the results mean and what are the norms for the test conducted?)

[Response Begins]

The multi-item measure with the highest Cronbach's alpha reliability (internal consistency) coefficient is Communication (alpha = 0.85; Table 2a.4). Four of the five multi-item measures meet the standard threshold of 0.70 or higher (Nunnally & Bernstein 1994), with the one exception, Help for Symptoms, nearing that threshold (alpha=0.69).

Item-total Pearson correlations are generally moderate to high (Table 2a.4). For all measures with 3 or more items, the Cronbach's alpha with item deleted is smaller than the alpha with all items. These results indicate that items are related to all other items within their multi-item measures.

Across the 28 programs that had at least 10 respondents, there is adequate variation in measure scores, as indicated by the ICC and reliability estimates shown in Table 2a.5. Six of the seven proposed measures exhibit acceptable program-level reliability of 0.70 or greater at 100 measure respondents; the remaining measure, Overall Rating, nears the threshold at reliability of 0.67 at 100 respondents.

[Response Ends]

2b.01. Select the level of validity testing that was conducted.

[Response Begins]

Patient or Encounter-Level (data element validity must address ALL critical data elements)

Accountable Entity Level (e.g. hospitals, clinicians)

Empirical validity testing

[Response Ends]

2b.02. For each level of testing checked above, describe the method of validity testing and what it tests.

Describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used.

[Response Begins]

For validity testing at the data element level, we used confirmatory factor analysis (CFA) to evaluate the factor structure of 18 evaluative survey items identified by our technical expert panel as those most important for assessing aspects of high-quality home-based serious illness care. We used means and variance adjusted weighted least squares to account for the dichotomous nature of top-box item scores (Flora and Curran, 2004). We used a criterion of factor loadings ≥ 0.40 for inclusion within the multi-item measure (Brown, 2015), and assessed overall model fit using the Comparative Fit Index (CFI), the root mean square error of approximation (RMSEA), and weighted root mean square residual (WRMR). Prior research indicates that a model with a good fit typically has a CFI > 0.95, RMSEA < 0.05, and WRMR < 1.0, with WRMR being less critical (Distefano et al., 2018; Hu and Bentler, 1999). The model chi-square statistic and standard error of model estimates were adjusted to account for the clustering of patients within programs (Wu and Kowk, 2012).

To assess construct validity at the quality measure level, we evaluated the associations of each multi-item measure's top-box score with the top-box score of the two global, single-item measures, Overall Rating of the Program and Willingness to Recommend the Program. We estimated multivariate linear regression models with the global measures as dependent variables to highlight the association of each composite with those measures. All models were adjusted for case mix and mode of survey administration (described below in 2b.20 and 2b.14), and were estimated with weighted least square mean and variance adjustment (WLSMV) in Mplus to correct for attenuation in regression coefficients with categorical outcomes (Muthén 1984). Standard errors and significance tests of regression coefficients were adjusted for clustering of patients within programs (Wu and Kowk, 2012). Missing data were handled using full-information likelihood estimation.

To assess the discriminant validity of the proposed multi-item measures at the quality measure level (i.e., degree to which the measures assess distinct content domains), we calculated correlations between the multi-item measure scores computed as the average of top-box-scored items, adjusting for clustering within programs. Correlations exceeding 0.80 may indicate that multi-item measures are measuring aspects of care that are insufficiently distinct (Brown 2015).

Citations:

Brown TA: Confirmatory Factor Analysis for Applied Research. 2nd ed. New York, NY: Guilford Press, 2015.

DiStefano C, Liu J, Jiang N, et al.: Examination of the weighted root mean square residual: Evidence for trustworthiness? Structural Equation Modeling: A Multidisciplinary Journal 2018;25(3):453-466.

Flora DB, Curran PJ: An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychol Methods 2004;9(4):466-491.

Hu Lt, Bentler PM: Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999;6(1):1-55.

Muthén B: A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. Psychometrika 1984;49(1):115-132.

Wu J-Y, Kwok O-m: Using SEM to analyze complex survey data: A comparison between design-based single-level and model-based multilevel approaches. Structural Equation Modeling: A Multidisciplinary Journal 2012;19(1):16-35.

[Response Ends]

2b.03. Provide the statistical results from validity testing.

Examples may include correlations or t-test results.

[Response Begins]

Table 2b.1. Factor Loadings of Component Items in Proposed Serious Illness Survey for Home-Based Programs Multi-Item Measures.

Composite and Global Measures & Component Survey Items	Factor Loading
Communication	-
In the last 3 months, how often did people from this program spend enough time with you when they visited?	0.84
In the last 3 months, how often did people from this program explain things to you in a way you could understand?	0.86
In the last 3 months, how often did people from this program listen carefully to you?	0.93
In the last 3 months, how often did you feel that people from this program cared about you as a whole person?	0.91
In the last 3 months, how often did you feel heard and understood by people from this program?	0.97
Care Coordination	-
In the last 3 months, how often did people from this program seem to know the important information about your medical history?	0.78
In the last 3 months, did someone from this program talk with you about the care or treatment you get from your other doctors or health care providers?	0.71
In the last 3 months, did someone from this program talk with you about all the medicines you are taking?	0.76

Composite and Global Measures & Component Survey Items	Factor Loading
Everyday activities include things like getting ready in the morning, getting meals, or going places in your community. In the last 3 months, did someone from this program talk with you about how to get help with everyday activities?	0.74
In the last 3 months, when you contacted this program between visits, did you get the help you needed?	0.80
Help for Symptoms	-
In the last 3 months, did you get as much help as you wanted for your pain?	0.83
In the last 3 months, did you get as much help as you wanted for your breathing?	0.78
In the last 3 months, did you get as much help as you wanted for your feelings of anxiety or sadness?	0.83
Planning for Care	-
Did someone from this program ever talk with you about what you should do during a health emergency?	0.85
Did someone from this program ever talk with you about what is important in your life?	0.85
Did someone from this program ever talk with you about what your health care options would be if you got sicker?	0.81
Support for Family and Friends	-
In the last 3 months, did the people from the program involve your family members or friends in discussions about your health care as much as you wanted?	0.82
In the last 3 months, did your family members or friends get as much emotional support as they wanted from this program?	0.98

Table shows factor loadings of component items in Serious Illness Survey for Home-Based Programs multi-item measures.

- This cell intentionally left empty.

Table 2b.2. Standardized Regression Coefficients from Using Multi-Item Serious Illness Survey for Home-Based Programs Measures to Predict Global Ratings.

Multi-Item Measure	Overall Rating of Program	Willingness to Recommend Program
Communication	0.56*	0.56*
Care Coordination	0.56*	0.57*
Help for Symptoms	0.46*	0.44*
Planning for Care	0.52*	0.50*
Support for Family and Friends	0.44*	0.47*

Table shows standardized regression coefficients from using multi-item measures to predict overall rating and willingness to recommend the home-based serious illness program.

*p<0.001

Table 2b.3. Correlations Among Proposed Serious Illness Care Multi-Item Measures.

Multi-Item Measure	Communication	Care Coordination	Help for Symptoms	Planning for Care	Support for Family and Friends
Communication	1	-	-	-	-
Care Coordination	0.62*	1	-	-	-
Help for Symptoms	0.40*	0.48*	1	-	-
Planning for Care	0.39*	0.58*	0.43*	1	-
Support for Family and Friends	0.44*	0.45*	0.43*	0.43*	1

Table shows correlation between Serious Illness Survey for Home-Based Programs multi-item measures.

* $p < 0.001$

- This cell intentionally left empty.

[Response Ends]

2b.04. Provide your interpretation of the results in terms of demonstrating validity. (i.e., what do the results mean and what are the norms for the test conducted?)

[Response Begins]

A five-factor CFA model provides an excellent fit to the data, $\chi^2_{(125)} = 269.45$; CFI = 0.992; RMSEA = 0.023; WRMR = 1.463. The factor loadings range between 0.71 to 0.98, suggesting that these data elements are strong indicators of the corresponding construct.

All five multi-item measures exhibit moderate to large associations with respondents' global assessments of serious illness program care, Overall Rating and Willingness to Recommend the Program (β=0.44 to β=0.57; Table 2b.2), indicating construct validity at the quality measure level. Care Coordination and Communication composites are the two strongest predictors of overall rating of care (β=0.56 and β=0.56, respectively) and willingness to recommend (β=0.57 and β=0.56, respectively).

The five multi-item measures are moderately correlated (Table 2b.3), suggesting that they measure related but distinct aspects of care (i.e., indicating discriminant validity at the quality measure level). Intercorrelations are highest between Care Coordination and other measures (r=0.45 to r=0.62), reflecting the core role that serious illness programs play in coordinating care through communication, planning, and assessing and managing symptoms.

[Response Ends]

2b.05. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified.

Describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided in Importance to Measure and Report: Gap in Care/Disparities.

[Response Begins]

We examined the ability to identify statistically significant and meaningful differences in measure scores across serious illness programs by calculating the number and percentage of programs that were significantly above or below the field test program average for each Serious Illness Survey for Home-Based Programs measure. Top-box

scores were adjusted for mode and case mix, and a two-sided alpha=0.05 level test was used to test for significance.

[Response Ends]

2b.06. Describe the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities.

Examples may include number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined.

[Response Begins]

Table 2b.1. Number of Serious Illness Programs Significantly Above or Below the Field Test Program Average

Measures	Count of Programs Significantly Above	Count of Programs Significantly Below	% Statistically Different from Field Test Program Average*
Communication	4	4	28.6%
Care Coordination	3	3	21.4%
Help for Symptoms	2	4	21.4%
Planning for Care	3	2	17.9%
Support for Family and Friends	2	4	21.4%
Overall Rating of Program	3	3	21.4%
Willingness to Recommend Program	3	5	28.6%

Table showing number of serious illness programs with measure scores significantly above and below the field test program average, and the percent of programs with scores statistically different from the field test program average.

[Response Ends]

2b.07. Provide your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities.

In other words, what do the results mean in terms of statistical and meaningful differences?

[Response Begins]

As shown in Table 2b.1, across measures from the Serious Illness Survey for Home-Based Programs, between 18 and 29 percent of programs participating in the field test scored either significantly above or below the field test program average. This indicates that the measures can identify statistically significant differences in programs' performance. Among program scores that were significantly above or below the average, the mean absolute difference between the programs' scores and the average program score for a given measure ranged from 11.0 for Communication to 17.9 for Care Planning, indicating large differences from the program average.

[Response Ends]

2b.08. Describe the method of testing conducted to identify the extent and distribution of missing data (or non-response) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and non-responders). Include how the specified handling of missing data minimizes bias.

Describe the steps—do not just name a method; what statistical analysis was used.

[Response Begins]

We assessed the association between survey nonresponse and several patient characteristics, including age, primary payer for care, and primary diagnosis.

In addition, we assessed nonresponse to evaluative items among unit respondents. We describe the proportion of respondents that skipped each item *appropriately* (i.e., dictated by the survey's skip logic instructions), *inappropriately* (i.e., *not* dictated by the survey's skip logic instructions), as well as the total proportion of missing data for each evaluative item on the survey.

[Response Ends]

2b.09. Provide the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data.

For example, provide results of sensitivity analysis of the effect of various rules for missing data/non-response. If no empirical sensitivity analysis was conducted, identify the approaches for handling missing data that were considered and benefits and drawbacks of each).

[Response Begins]

The overall field test response rate was 36.4%, with response rates differing significantly by mode of administration (30.4% in mail-only mode and a 42.5% in mail-telephone mode]). Overall, older patients were more likely to respond than younger patients, and response rates increased with age (DeYoreo et al., 2022). Those with cancer were less likely to respond (OR: 0.83) compared to those with other diagnoses. Individuals with Medicaid as the primary payer for care were less likely to respond (OR: 0.69) than those with Medicare as the primary payer. The probability of response was higher for those with more recent visits and for those with more in-person visits. Patients with more visits had higher response rates than those with fewer visits ($p<0.01$).

Table 2b.2. Serious Illness Survey for Home-Based Programs Missing Data

Measures & Component Survey Items	% Missing due to Appropriate Skip[a]	% Missing due to Inappropriate Skip	% Missing (Total)
Communication	-	-	-
In the last 3 months, how often did people from this program spend enough time with you when they visited? (Q3)	0.0	3.9	3.9
In the last 3 months, how often did people from this program explain things to you in a way you could understand? (Q4)	0.0	4.1	4.1
In the last 3 months, how often did people from this program listen carefully to you? (Q5)	0.0	4.3	4.3

Measures & Component Survey Items	% Missing due to Appropriate Skip[a]	% Missing due to Inappropriate Skip	% Missing (Total)
In the last 3 months, how often did you feel that people from this program cared about you as a whole person? (Q6)	0.0	4.7	4.7
In the last 3 months, how often did you feel heard and understood by people from this program? (Q7)	0.0	4.1	4.1
Care Coordination	-	-	-
In the last 3 months, how often did people from this program seem to know the important information about your medical history? (Q8)	0.0	4.0	4.0
In the last 3 months, did someone from this program talk with you about the care or treatment you get from your other doctors or health care providers? (Q9)	0.0	4.9	4.9
In the last 3 months, did someone from this program talk with you about all the medicines you are taking? (Q10)	0.9	4.0	4.9
Everyday activities include things like getting ready in the morning, getting meals, or going places in your community. In the last 3 months, did someone from this program talk with you about how to get help with everyday activities? (Q11)	18.4	6.1	24.4
In the last 3 months, when you contacted this program between visits, did you get the help you needed? (Q12)	63.5	1.1	64.6
Help for Symptoms	-	-	-
In the last 3 months, did you get as much help as you wanted for your pain? (Q14)	44.2	2.2	46.4
In the last 3 months, did you get as much help as you wanted for your breathing? (Q16)	68.4	0.9	69.3
In the last 3 months, did you get as much help as you wanted for your feelings of anxiety or sadness? (Q18)	63.9	1.4	65.2
Planning for Care	-	-	-
Did someone from this program ever talk with you about what you should do during a health emergency? (Q22)	0.0	5.9	5.9
Did someone from this program ever talk with you about what is important in your life? (Q23)	0.0	5.9	5.9
Did someone from this program ever talk with you about what your health care options would be if you got sicker? (Q24)	0.0	5.7	5.7
Support for Family and Friends	-	-	-

Measures & Component Survey Items	% Missing due to Appropriate Skip[a]	% Missing due to Inappropriate Skip	% Missing (Total)
In the last 3 months, did the people from the program involve your family members or friends in discussions about your health care as much as you wanted? (Q20)	23.3	2.7	26.0
In the last 3 months, did your family members or friends get as much emotional support as they wanted from this program? (Q21)	39.3	4.4	43.7
Overall Rating of the Program	-	-	-
Using any number from 0 to 10, where 0 is the worst care possible and 10 is the best care possible, what number would you use to rate your care from this program? (Q25)	0.0	6.4	6.4
Willingness to Recommend the Program	-	-	-
Would you recommend this program to your friends and family? (Q26)	0.0	6.0	6.0

Table showing the percent of survey responses missing due to appropriate and inappropriate skipping of items by respondents to the Serious Illness Survey for Home-Based Programs.

[a] Appropriate skips include missingness due skip logic, as well as selection of tailored non-applicable response options (e.g., "I did not want help for my pain.")

- This cell intentionally left empty.

Citation:

DeYoreo M, Anhang Price R, Bradley MA, Schlang D, Montemayor CK, Tolpadi A, Cleary PD, Teno JM, and Elliott MN. Adding telephone follow-up can improve representativeness of surveys of seriously ill people. *J Am Geratr Soc.* 2022 Jun;70(6):1870-1873.

[Response Ends]

2b.10. Provide your interpretation of the results, in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and non-responders), and how the specified handling of missing data minimizes bias.

In other words, what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis was conducted, justify the selected approach for missing data.

[Response Begins]

Mail only response rates to the field test of the Serious Illness Survey for Home-Based Programs were similar to those observed for other surveys of care experiences in national use, and mail-telephone response rates to the field test were somewhat higher than those observed in such national surveys. For example, national response rates to the CAHPS Hospice Survey, which is administered predominantly in mail-only mode, have been approximately 30 percent in recent years (CAHPS Hospice Survey website); response rates to the Medicare Advantage, Fee-for-Service, and Prescription Drug Plan Surveys, which are conducted using mail-telephone mode, were 32 to 39 percent in 2019 (Medicare Advantage and Prescription Drug Plan CAHPS Survey website). Propensity to respond to the Serious Illness Survey for Home-Based Programs varies with patient characteristics in ways that

are similar to those observed for other patient surveys. To address these differences, we recommend the use of statistical methods to avoid nonresponse bias, which occurs when those who respond tend to provide systematically higher or lower evaluations than those who do not respond. Use of nonresponse weights to eliminate nonresponse bias increases the variance of estimates and does not necessarily improve the precision of between-program comparisons; however, case-mix adjustment often eliminates most of the nonresponse bias and allows for comparison of programs that differ in terms of patient composition (Elliott, Edwards, Angeles, et al. 2005). We therefore recommend an approach consistent with that taken in Centers for Medicare & Medicaid Services' CAHPS initiatives that program scores be adjusted for case mix in order to allow for fair comparison across programs.

Across evaluative items, less than 6.5 percent of respondents inappropriately skipped items. Item missingness tended to be higher toward the end of the survey. These findings suggest that it is unlikely that Serious Illness Survey for Home-Based Program results are biased due to systematic skipping of items by respondents.

Citation:

CAHPS Hospice Survey official website, Scoring and Analysis page. https://hospicecahpssurvey.org/en/public-reporting/scoring-and-analysis/. Last accessed July 27, 2022.

Elliott MN, Edwards C, Angeles J, Hays RD (2005). "Patterns of unit and item non-response in the CAHPS® Hospital Survey." *Hlth Serv Res* 40(6): 2096-2119.

Medicare Advantage and Prescription Drug Plan CAHPS Survey official website, Comparative Data page. https://www.ma-pdpcahps.org/en/comparative-data/. Last accessed July 27, 2022.

[Response Ends]

Note: This item is directed to measures that are risk-adjusted (with or without social risk factors) OR to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eCQMs). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.

2b.11. Indicate whether there is more than one set of specifications for this measure.

[Response Begins]

Yes, there is more than one set of specifications for this measure

[Response Ends]

2b.12. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications.

Describe the steps—do not just name a method. Indicate what statistical analysis was used.

[Response Begins]

As described in sp.27, the Serious Illness Survey for Home-Based Programs field test was conducted in both mail only and mail-telephone modes; these modes are also described and recommended in the online guidance for administration of the survey available at: https://www.rand.org/health-care/surveys_tools/serious-illness-

survey.html. We used linear regression models to estimate the effect of mode of administration on survey measure scores.

[Response Ends]

2b.13. Provide the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications.

Examples may include correlation, and/or rank order.

[Response Begins]

Mixed mode of survey administration is associated with slightly more positive assessments than mail-only administration for all measures other than Overall Rating and Willingness to Recommend the Program, although these associations do not meet thresholds for statistical significance (DeYoreo et al., 2022).

Citation:

DeYoreo M, Anhang Price R, Montemayor CK, Tolpadi A, Bradley MA, Schlang D, Teno JM, Cleary PD, and Elliott MN. 2022. Adjusting for patient characteristics to compare quality of care provided by serious illness programs. *J Pall Med.* 25(7).

[Response Ends]

2b.14. Provide your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications.

In other words, what do the results mean and what are the norms for the test conducted.

[Response Begins]

Survey mode is generally a choice that affects a program's entire sample (in contrast to the field test, in which mode was randomized at the patient level within programs). Therefore, survey mode is expected to have a somewhat larger impact on program scores than is suggested by the patient-level results described in 2b.13. Therefore, we recommend the conservative approach of adjusting for mode of survey administration when calculating Serious Illness Survey for Home-Based Programs measure scores.

[Response Ends]

2b.15. Indicate whether the measure uses exclusions.

[Response Begins]

Yes, the measure uses exclusions.

[Response Ends]

2b.16. Describe the method of testing exclusions and what was tested.

Describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used?

[Response Begins]

As noted in sp.17, the Serious Illness Survey for Home-Based Programs is designed for administration to adult patients who are currently enrolled in home-based serious illness programs. Patients are excluded from the survey sample if they:

- Are under age 18
- Receive care from a serious illness program in a setting OTHER than home or an assisted living facility (e.g., in a nursing home or other long-term care facility)
- Are known to have been discharged to hospice
- Are known to have died
- Have been enrolled for in the serious illness program for less than six weeks as of the date of survey sampling

Eligibility criteria and exclusions for the Serious Illness Survey for Home-Based Programs are designed to ensure that only those who are currently receiving care from home-based serious illness programs, and who have sufficient experience with such a program, are included in the sample. Exclusions are not based on statistical testing.

Proxy responses are permitted to allow for inclusion of care assessments for patients whose functional status or cognitive ability may make it difficult to complete the survey independently.

[Response Ends]

2b.17. Provide the statistical results from testing exclusions.

Include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores.

[Response Begins]

N/A

[Response Ends]

2b.18. Provide your interpretation of the results, in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results.

In other words, the value outweighs the burden of increased data collection and analysis. Note: If patient preference is an exclusion, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion.

[Response Begins]

N/A

[Response Ends]

2b.19. Check all methods used to address risk factors.

[Response Begins]

Statistical risk model with risk factors (specify number of risk factors)

[Statistical risk model with risk factors (specify number of risk factors) Please Explain]

8; section 2.20 lists all of the proposed risk factors in detail.

[Response Ends]

2b.20. If using statistical risk models, provide detailed risk model specifications, including the risk model method, risk factors, risk factor data sources, coefficients, equations, codes with descriptors, and definitions.

[Response Begins]

The recommended case-mix variables for adjustment of Serious Illness Survey for Home-Based Programs measures are:

- Age (categories: 18–54, 55–64, 65–69, 70–74, 75–79, 80–84, 85–89, 90 or older)
- Education (categories: 8th grade or less, some high school, high school graduate, some college, 4-year college graduate, more than 4-year college)
- Primary diagnosis (categories: cancer, Alzheimer's and other dementias, other)
- Proxy response (categories: proxy answered, proxy helped in some other way, no proxy)
- Self-reported functional status (categories: not able to get out of bed, able to get out of bed but not house, able to leave house)
- Self-reported physical health (categories: excellent, very good, good, fair, poor)
- Self-reported mental health (categories: excellent, very good, good, fair, poor)
- Response percentile (a continuous, within-program rank-based measure of the time between survey administration and survey response, calculated as the rank-ordered number of days between the first day of survey administration and survey response for respondents in each program and mode, relative to all eligible patients within each program and mode, scaled from 0 to 1)

To calculate case-mix adjusted Serious Illness Survey for Home-Based Programs measure scores, users can fit a linear model for each survey item (e.g., with each survey item as the response variable) that includes as covariates each case-mix adjustor, mode of survey administration, and fixed effects for entities to be compared (e.g., program fixed effects when comparing programs).

The models described above are linear models which can be expressed as follows:

$$Y_{ij} = \alpha + \gamma_j + \sum_k \chi_{ijk}\beta_k + \varepsilon_{ij}$$

Where Y subscript ij is the rating given by patient i in program j, and X subscript ij is a vector of covariates for this patient (age, education, primary diagnosis, etc., coded as dummy variables given the categories described above), gamma subscript j is a program fixed effect, and epsilon subscript ij is an error term with mean 0 and constant variance.

For each item, the fitted linear model can be used to generate adjusted scores for each program (i.e., generate predicted scores that assume that each program has the same case mix and mode of survey administration). For multi-item measures, average the adjusted top-box scores for the items that compose the measure.

[Response Ends]

2b.21. If an outcome or resource use measure is not risk-adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (i.e., case mix) is not needed to achieve fair comparisons across measured entities.

[Response Begins]

[Response Ends]

2b.22. Select all applicable resources and methods used to develop the conceptual model of how social risk impacts this outcome.

[Response Begins]

Published literature

Internal data analysis

Other (specify)

[Other (specify) Please Explain]

Input from technical experts

[Response Ends]

2b.23. Describe the conceptual and statistical methods and criteria used to test and select patient-level risk factors (e.g., clinical factors, social risk factors) used in the statistical risk model or for stratification by risk.

Please be sure to address the following: potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10 or other statistical tests; correlation of x or higher. Patient factors should be present at the start of care, if applicable. Also discuss any "ordering" of risk factor inclusion; note whether social risk factors are added after all clinical factors. Discuss any considerations regarding data sources (e.g., availability, specificity).

[Response Begins]

Patient characteristics like age and education can be related to how patients respond to survey questions (Elliott et al., 2009; Paddison et al., 2012). Since there is great variability across serious illness programs with regard to patient mix, program-level scores can be compared fairly only after adjusting for patient characteristics that influence how patients respond to questions about quality of care but are not under the control of the programs. We sought to identify patient characteristics that are exogenous to the care provided by the program and both strongly predict survey measure scores and impact program scores for at least one outcome measure. Patient characteristics considered as potential candidates for case-mix adjustment included sex, age, primary diagnosis, education, language spoken at home, and self-reported ratings of mental health, physical health, and functional status (DeYoreo et al., 2022). In keeping with adjustments for other patient experience surveys, we also considered whether a proxy assisted with completion of the survey on behalf of the patient (which may capture severity of illness and/or cognitive impairments; Elliott et al., 2008), and ''response percentile,'' defined as the rank-ordered time between initiation of survey administration and response for each respondent relative to all eligible patients within program and mode, scaled from 0 to 1 (which captures both the program response rate and how soon a person responded compared with others in the same program and mode; Zaslavsky et al., 2002).

We used linear regression models to estimate the effect of potential case-mix adjustors on survey measure scores and assessed whether the regression coefficient associated with the adjustor was statistically significantly different from zero. We also evaluated the impact of each adjustor on program-level scores by comparing the program-level scores with and without the adjustor of interest. We first fit regression models where the outcomes were the measure scores and the predictors included all candidate case-mix adjustor variables, mode of survey administration (mail-only vs. mixed mode), and program fixed effects. The regression coefficients can be interpreted as the average effect of each patient characteristic on outcomes within a program. We identified the adjustors that were statistically significantly predictive of at least one outcome ($p < 0.01$) and interpreted the effect of each adjustor on assessments. Models were restricted to programs with 10 or more respondents (28 out of 32 programs) to limit attenuation bias.

To evaluate the impact of each case-mix adjustor on adjustments, we fit a series of models that removed one candidate adjustor at a time. For each composite score and global rating item, we calculated the correlation between the adjusted program-level scores from the full models and the adjusted program-level scores from the

models that left out the case-mix adjustor variable of interest, assuming each program had population-average case mix and mode of survey administration. Adjusted program scores were generated for each item using the estimated regression coefficients and the characteristics of respondents in the program. The quantity 1 - r^2 represents the proportion of variance in the adjusted scores marginally associated with adjustment for that variable and indexes each adjustor's marginal impact on program-level scores. Only characteristics that vary among programs and predict patient responses affect program-level scores. Our recommended set of case-mix adjustors includes all variables that were statistically significantly associated with respondent evaluations ($p < 0.01$) and which have an impact (measured by 1 - r^2) of at least 1 percent for one or more outcome measures.

To determine the final case-mix adjustment model, we also incorporated feedback from our team's expert advisors regarding variables that did not meet these empirical criteria, but which are important for face validity. For example, if people think that patients with poor function tend to give lower ratings and/or tend to report negative experiences, programs with many such patients may think they are being unfairly evaluated unless there is an adjustment for the proportion of such patients in each program, irrespective of empirical evidence of impact. Once we identified an appropriate CMA model, we determined the overall impact of adjustment on program scores by comparing case-mix adjusted program-level scores to those not adjusted for case-mix.

Citations:

DeYoreo M, Anhang Price R, Montemayor CK, Tolpadi A, Bradley MA, Schlang D, Teno JM, Cleary PD, and Elliott MN. 2022. Adjusting for patient characteristics to compare quality of care provided by serious illness programs. *J Pall Med.* 25(7).

Elliott MN, Beckett MK, Chong K, et al.: How do proxy responses and proxy-assisted responses differ from what Medicare beneficiaries might have reported about their health care? Health Services Research 2008;43(3):833-848.

Elliott MN, Zaslavsky AM, Goldstein E, et al.: Effects of survey mode, patient mix, and nonresponse on CAHPS® hospital survey scores. Health Services Research 2009;44(2 Pt 1):501-518.

Paddison C, Elliott M, Parker R, et al.: Should measures of patient experience in primary care be adjusted for case mix? Evidence from the UK General Practice Patient Survey. BMJ Quality & Safety 2012;21(8):634-640.

Zaslavsky AM, Zaborski LB, Cleary PD: Factors affecting response rates to the Consumer Assessment of Health Plans Study survey. Medical Care 2002;40(6):485-499.

[Response Ends]

2b.24. Detail the statistical results of the analyses used to test and select risk factors for inclusion in or exclusion from the risk model/stratification.

[Response Begins]

Table 2b.3 provides the regression coefficient estimates and standard errors from a multivariate model that includes all of the proposed candidate adjustors except for sex, which did not exhibit a statistically significant relationship with any of the measure scores at the 0.01 significance level. Directions of association are generally similar to what has been seen previously in the patient experience literature (Cefalu et al. 2021). Response percentile is significantly negatively associated with assessments for two of seven outcomes (Communication and Care Coordination). Education is significantly associated with Overall Rating (negatively); age is significantly associated with two outcomes (negatively). Diagnosis is significantly associated with two outcomes (Help for Symptoms and Planning for Care). For these outcomes, compared to all other diagnoses, cancer diagnosis is associated with significantly more positive assessments and Alzheimer's and other dementias is associated with significantly more positive assessments for Help for Symptoms. Proxy use is significant for four outcomes, with both proxy response and other help from a proxy being associated with more positive responses compared to no proxy. Those with poor function tended to respond more negatively (though functional status was only significant for Care Coordination). Self-rated physical health is not significantly associated with any measure. Self-rated mental health is significantly associated with five of the measures; better ratings of mental health were generally associated with more positive assessments.

Table 2b.3. Relationship between Case-mix Adjustors and Serious Illness Survey for Home-Based Programs Measures.

-	**Communication** β (SE)	**Care Coordination** β (SE)	**Help for Symptoms** β (SE)	**Planning for Care** β (SE)	**Support for Family and Friends** β (SE)	**Overall Rating** β (SE)	**Willingness to Recommend** β (SE)
Education	-	-	-	-	-	**	-
Education: 8th grade or less	0.01 (0.03)	0.02 (0.03)	-0.04 (0.04)	0.04 (0.03)	0.11 (0.04) **	0.01 (0.04)	0.03 (0.04)
Education: Some high school but did not graduate	0.00 (0.03)	0.04 (0.02)	0.03 (0.04)	0.05 (0.03)	0.03 (0.04)	0.05 (0.03)	0.02 (0.03)
Education: High school graduate or GED (ref)	-	-	-	-	-	-	-
Education: Some college or 2-year degree	0.00 (0.02)	0.00 (0.02)	-0.02 (0.03)	-0.03 (0.02)	-0.02 (0.03)	-0.06 (0.03)*	0.00 (0.03)
Education: 4-year college graduate	0.00 (0.03)	0.03 (0.03)	0.05 (0.04)	0.01 (0.03)	0.03 (0.04)	0.00 (0.04)	0.05 (0.04)
Education: More than 4-year college degree	0.00 (0.03)	-0.05 (0.03)	-0.06 (0.04)	-0.03 (0.03)	0.03 (0.04)	-0.09 (0.04)	0.00 (0.04)
Age	-	**	-	**	-	-	*
Age: 18-54	0.11 (0.04) **	0.15 (0.04) **	0.12 (0.06)	0.12 (0.05) *	0.07 (0.06)	0.09 (0.06)	0.17 (0.06) **
Age: 55-64	0.04 (0.03)	0.10 (0.03) **	0.05 (0.05)	0.17 (0.04) **	0.01 (0.05)	0.01 (0.04)	0.03 (0.04)
Age: 65-69	0.05 (0.03)	0.11 (0.03) **	0.03 (0.05)	0.11 (0.04) **	-0.04 (0.05)	0.03 (0.04)	0.06 (0.04)
Age: 70-74	0.05 (0.03)	0.07 (0.03) *	0.04 (0.04)	0.11 (0.03) **	0.02 (0.04)	0.04 (0.04)	0.10 (0.04) **
Age: 75-79	0.04 (0.03)	0.05 (0.03)	0.00 (0.04)	0.04 (0.03)	-0.03 (0.04)	0.04 (0.04)	0.02 (0.04)
Age: 80-84	0.02 (0.02)	0.06 (0.02) **	0.01 (0.04)	0.07 (0.03)*	-0.02 (0.03)	0.04 (0.03)	0.05 (0.03)

-	**Communication** β (SE)	**Care Coordination** β (SE)	**Help for Symptoms** β (SE)	**Planning for Care** β (SE)	**Support for Family and Friends** β (SE)	**Overall Rating** β (SE)	**Willingness to Recommend** β (SE)
Age: 85-89	0.02 (0.02)	0.01 (0.02)	-0.03 (0.04)	0.01 (0.03)	-0.03 (0.03)	0.01 (0.03)	0.01 (0.03)
Age: >=90 (ref)	-	-	-	-	-	-	-
Diagnosis	*	-	**	**	-	-	-
Diagnosis: Cancer	0.05 (0.02)*	0.03 (0.02)	0.18 (0.03) **	0.09 (0.03) **	0.05 (0.03)	0.06 (0.03)	0.04 (0.03)
Diagnosis: Alzheimer's & Other Dementias	0.05 (0.03)	0.03 (0.03)	0.12 (0.05)**	0.01 (0.03)	-0.01 (0.04)	-0.02 (0.04)	0.04 (0.04)
Diagnosis: Other (ref)	-	-	-	-	-	-	-
Language Spoken at Home	-	-	-	*	-	-	-
Language: English (ref)	-	-	-	-	-	-	-
Language: Spanish	-0.04 (0.04)	0.06 (0.04)	0.10 (0.06)	0.13 (0.05) **	0.02 (0.06)	0.08 (0.05)	0.08 (0.05)
Language: Other	-0.05 (0.04)	0.01 (0.04)	0.07 (0.07)	-0.03 (0.05)	-0.01 (0.07)	-0.03 (0.06)	0.00 (0.06)
Proxy Assistance with Survey Response	**	**	*	**	**	-	-
Proxy assistance: Proxy answered questions	0.05 (0.02) **	0.05 (0.02) **	0.09 (0.03) **	0.09 (0.02) **	0.21 (0.03) **	0.03 (0.03)	0.00 (0.03)
Proxy assistance: Proxy helped in some other way	0.08 (0.02) **	0.09 (0.02) **	0.04 (0.04)	0.09 (0.03) **	0.20 (0.03) **	0.01 (0.03)	0.02 (0.03)

-	**Communication** β (SE)	**Care Coordination** β (SE)	**Help for Symptoms** β (SE)	**Planning for Care** β (SE)	**Support for Family and Friends** β (SE)	**Overall Rating** β (SE)	**Willingness to Recommend** β (SE)
Proxy assistance: No proxy (ref)	-	-	-	-	-	-	-
Self-Reported Functional status	*	**	-	-	-	-	-
Functional status: Not able to get out of bed	-0.08 (0.03) *	-0.09 (0.03) **	0.00 (0.05)	-0.07 (0.04)	-0.03 (0.04)	-0.09 (0.04) *	-0.02 (0.04)
Functional status: Can get out of bed but not leave house	-0.04 (0.03)	-0.06 (0.03) *	-0.06 (0.05)	-0.03 (0.03)	-0.02 (0.04)	-0.03 (0.04)	-0.04 (0.04)
Functional status: Able to leave house (ref)	-	-	-	-	-	-	-
Self-Reported Physical Health	-	-	-	-	-	-	-
Physical Health: Excellent	-0.05 (0.05)	0.00 (0.05)	-0.03 (0.09)	-0.03 (0.06)	-0.07 (0.08)	-0.07 (0.07)	-0.13 (0.07)
Physical Health: Very good	0.01 (0.03)	0.04 (0.03)	0.01 (0.05)	-0.02 (0.03)	0.00 (0.04)	0.05 (0.04)	0.00 (0.04)
Physical Health: Good (ref)	-	-	-	-	-	-	-
Physical Health: Fair	0.01 (0.02)	0.01 (0.02)	-0.03 (0.03)	0.00 (0.02)	-0.02 (0.03)	-0.01 (0.03)	0.01 (0.03)
Physical Health: Poor	-0.01 (0.02)	0.00 (0.02)	0.00 (0.04)	0.03 (0.03)	-0.02 (0.03)	-0.03 (0.03)	0.02 (0.03)

-	**Communication** β (SE)	**Care Coordination** β (SE)	**Help for Symptoms** β (SE)	**Planning for Care** β (SE)	**Support for Family and Friends** β (SE)	**Overall Rating** β (SE)	**Willingness to Recommend** β (SE)
Self-Reported Mental Health	**	**	**	**	-	**	*
Mental health: Excellent	0.05 (0.03) *	0.09 (0.03) **	0.11 (0.04) *	0.10 (0.03) **	0.09 (0.04) *	0.09 (0.04) *	0.08 (0.04) *
Mental health: Very good	0.00 (0.02)	0 .00(0.02)	0.03 (0.03)	0.01 (0.02)	0.01 (0.03)	0.00 (0.03)	0.00 (0.03)
Mental health: Good (ref)	-	-	-	-	-	-	-
Mental health: Fair	-0.06 (0.02) **	-0.08 (0.02) **	-0.09 (0.03) **	-0.06 (0.02) *	-0.03 (0.03)	-0.08 (0.03) **	-0.05 (0.03)
Mental health: Poor	-0.05 (0.03)	-0.07 (0.03) *	-0.13 (0.05) **	-0.05 (0.04)	-0.03 (0.04)	-0.03 (0.04)	-0.02 (0.04)
Response Percentile[a]	-0.21 (0.06) **	-0.15 (0.06) **	-0.01 (0.09)	0.03 (0.07)	-0.17 (0.08) *	-0.11 (0.08)	0.00 (0.08)

Table shows estimates and standard errors from a model predicting Serious Illness Survey for Home-Based Programs measure scores with proposed candidate adjustors.

* p<0.05; **p<0.01

NOTE: P-values from joint tests of significance (F-tests) for the adjustor are provided as well as P-values that compare each level of the adjustor to the reference level.

- This cell intentionally left empty.

Table 2b.4 contains the impact measure values for the recommended set of case-mix adjustment variables, with larger values indicating that an adjustor plays a larger role. Values indicate the proportion of variance in patient-level scores that is uniquely associated with a given predictor, with three percent indicating a moderate impact (impact values exceed three percent for age on two measures, diagnosis on one measure) and one percent indicating a small impact. Age is the most important adjustor overall, followed by diagnosis.

Table 2b.4. Impact of each Case-mix Adjustor on Serious Illness Survey for Home-Based Programs Program-Level Measure Scores, in terms of percent of adjustment attributable to each variable.

Variable Removed	**Communication** %	**Care Coordination** %	**Help for Symptoms** %	**Planning for Care** %	**Support for Family and Friends** %	**Overall Rating of the Program** %	**Willingness to Recommend the Program** %	**Mean Impact** %
Age	1.0	3.2	2.3	3.4	0.5	0.4	1.3	1.7

Variable Removed	Communication %	Care Coordination %	Help for Symptoms %	Planning for Care %	Support for Family and Friends %	Overall Rating of the Program %	Willingness to Recommend the Program %	Mean Impact %
Diagnosis	0.7	0.2	3.9	1.3	0.5	0.6	0.3	1.1
Proxy Respondent	0.6	0.6	1.2	0.8	2.3	0.1	0	0.8
Self-rated mental health	0.4	0.6	1.8	0.4	0.6	0.7	0.2	0.7
Response percentile	1.5	0.6	0.2	0	0.7	0.3	0	0.5
Education	0	0.3	0.3	0.4	0.7	1.2	0	0.4
Functional status	0.2	0.5	0.2	0.1	0	0.3	0.1	0.2
Self-rated physical health	0.1	0.1	0.2	0	0	0.2	0.1	0.1

Table shows the impact of each case-mix adjustment variable on Serious Illness Survey for Home-Based Programs measure scores, in terms of the percent adjustment attributable to each variable, and providing a mean percent impact across measures.

Citation:

Cefalu M, Elliott MN, Hays RD: Adjustment of patient experience surveys for how people respond. Med Care 2021;59(3):202-205.

[Response Ends]

2b.25. Describe the analyses and interpretation resulting in the decision to select or not select social risk factors.

Examples may include prevalence of the factor across measured entities, availability of the data source, empirical association with the outcome, contribution of unique variation in the outcome, or assessment of between-unit effects and within-unit effects. Also describe the impact of adjusting for risk (or making no adjustment) on providers at high or low extremes of risk.

[Response Begins]

To determine which social risk factors should be included in the case-mix adjustment model, we screened the three social risk factors available in the data (language, education, and payer) for data quality; those with sufficient data quality were further screened for their variation in the sample and meaningful impact on measure scores. Payer data were missing for approximately one-fifth of respondents; therefore, payer was not included in further assessments.

As with all other candidate variables for the case-mix adjustment model, we applied criteria of a statistically significant association with at least one quality measure score at the 0.01 level and 1 - r^2 of at least 1 percent for at least one quality measure score. As shown in Tables 2ba.3 and 2ba.4, education was significantly, negatively associated with overall rating of care from the program at $p<0.01$ and met the 1-r^2 threshold. Language was

borderline significantly associated with one outcome (p = 0.01 for Planning for Care, with Spanish language associated with more negative assessments). As language did not meet the criteria for inclusion as a case-mix adjustor, we excluded language from the final recommended adjustment model.

[Response Ends]

2b.26. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (describe the steps—do not just name a method; what statistical analysis was used). Provide the statistical results from testing the approach to control for differences in patient characteristics (i.e., case mix) below. If stratified ONLY, enter "N/A" for questions about the statistical risk model discrimination and calibration statistics.

Validation testing should be conducted in a data set that is separate from the one used to develop the model.

[Response Begins]

To investigate the overall effect of case-mix adjustment on each Serious Illness Survey for Home-Based Programs measure, we compare program scores with and without adjustment (Table 2b.5). We additionally report Kendall's tau, t, a rank-based measure of correlation (-1 to +1) that can be used to calculate p = (1 - t)/2, which can be interpreted as the chance of mistakenly ranking one program as better than the other without adjustment, or the proportion of pairs of programs whose relative rankings are reversed by adjustment. Thus, t = 1 indicates that adjustment has no effect on relative rankings, and t = 0.8 indicates that there is a 10 percent chance that one program will be mistakenly ranked as higher than another (without adjustment). Analyses were restricted to programs with 10 or more respondents (28 out of 32 programs). Fully adjusted program scores are constructed from models which adjust for the final set of recommended case-mix adjustors, survey mode and include program fixed effects. Unadjusted program scores are constructed from models which adjust for survey mode and include program fixed effects.

Table 2b.5. Overall Impact of Adjustment.

Measure	1-r^2 between unadjusted and fully adjusted scores	Kendall's Tau	Percent of Program Pairs That Would Switch Rankings
Communication	0.05	0.88	6
Care Coordination	0.06	0.86	7
Help for Symptoms	0.10	0.76	12
Planning for Care	0.08	0.81	10
Support for Family and Friends	0.06	0.85	8
Overall Rating of the Program	0.05	0.89	6
Willingness to Recommend the Program	0.03	0.93	4

Table shows the overall impact of case-mix adjustment on measure scores, using Kendall's tau and displaying the percent of pairs of programs that switch rankings in the absence of adjustment.

[Response Ends]

2b.27. Provide risk model discrimination statistics.

For example, provide c-statistics or R-squared values.

[Response Begins]

The R-squared values from the fitted regression model with all case-mix adjustors, and fixed effects for program and mode of survey administration are shown in Table 2b.6. The measures with the highest R-squared are Help for Symptoms, Planning for Care, and Support for Family and Friends. The R-squared values are relatively low, which is expected in patient-level modeling due to patient-level variability. However, case-mix adjustment does have important effects for programs with very different case-mix distributions, as shown above in 2b.26, which illustrates the percent of program pairs that would switch rankings due to case-mix adjustment.

Table 2b.6. Model Discrimination Statistics from a Fitted Regression Model Predicting Serious Illness Survey for Home-Based Programs Measure Scores with Case-Mix Adjustors.

Measure	R-squared
Communication	0.07
Care Coordination	0.09
Help for Symptoms	0.11
Planning for Care	0.11
Support for Family and Friends	0.12
Overall Rating of the Program	0.06
Willingness to Recommend the Program	0.07

Table shows R-squared values from a fitted regression model predicting Serious Illness Survey for Home-Based Programs measure scores from all case-mix adjustors.

[Response Ends]

2b.28. Provide the statistical risk model calibration statistics (e.g., Hosmer-Lemeshow statistic).

[Response Begins]

Table 2b.7 presents the correlation between the observed and model-predicted (fitted) values, for each measure. Note that such results are consistent with what would be expected from patient experience survey responses (in contrast to a clinical outcome, for which a risk model often explains a large portion of the variation). Our case-mix adjustment model accounts for patient-level factors that are likely to matter based on the evidence from similar surveys and from the literature, and thus is accounting for flaws in measurement equivalence that would be present with unadjusted scores.

Table 2b.7. Correlation between Observed and Model-Predicted Values for Serious Illness Survey for Home-Based Programs Measures.

Measure	Correlation
Communication	0.27
Care Coordination	0.30
Help for Symptoms	0.33
Planning for Care	0.33
Support for Family and Friends	0.35
Overall Rating of the Program	0.25
Willingness to Recommend the Program	0.26

Table shows correlation between observed and predicted values for each Serious Illness Survey for Home-Based Programs measure.

[Response Ends]

2b.29. Provide the risk decile plots or calibration curves used in calibrating the statistical risk model.

The preferred file format is .png, but most image formats are acceptable.

[Response Begins]

Because we use top-box scoring, typical goodness of fit plots that compare observed, predicted, and residuals are not useful. Additionally, as noted in our prior response, for a patient experience survey, we do not expect the case-mix adjustment model to predict a great deal of the variation in the response. Importantly, we adjust for patient-level risk factors that are expected to matter based on the literature, and case-mix adjustment does have an important impact on programs that serve a very different patient mix.

[Response Ends]

2b.30. Provide the results of the risk stratification analysis.

[Response Begins]

N/A

[Response Ends]

2b.31. Provide your interpretation of the results, in terms of demonstrating adequacy of controlling for differences in patient characteristics (i.e., case mix).

In other words, what do the results mean and what are the norms for the test conducted?

[Response Begins]

Results identify patient-level variables that are significantly associated with respondent evaluations within programs and have a meaningful impact on program-level scores for at least one outcome measure. The adjustors that have the most impact on program-level scores are age, primary diagnosis, and proxy respondent. To ensure that comparison of program scores is fair across serious illness programs, scores should be adjusted for patient-level variables, which influence patients' reports about care quality, but are not under the control of the program administering care.

Table 2b.5 summarizes the impact of adjustment with all variables proposed for the case-mix adjustment model (''full adjustment'') on each outcome of interest. Kendall's tau comparing scores between the null and fully adjusted model for each measure are in the expected range, from 0.76 for Help for Symptoms to 0.93 for Willingness to Recommend the Program. Full adjustment has the most impact on Help for Symptoms, with 12% of pairs of programs having their relative rankings reversed by adjustment and 10% of the variance in program scores attributable to extraneous factors controlled for by the case-mix adjustment model.

[Response Ends]

2b.32. Describe any additional testing conducted to justify the risk adjustment approach used in specifying the measure.

Not required but would provide additional support of adequacy of the risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed.

[Response Begins]

None.

[Response Ends]

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3.01. Check all methods below that are used to generate the data elements needed to compute the measure score.

[Response Begins]

Other (Please describe)

[Other (Please describe) Please Explain]

Patient-reported data are collected via a survey instrument. Patient eligibility is determined using administrative data from the home-based serious illness program; this information may be available in the electronic health record.

[Response Ends]

3.02. Detail to what extent the specified data elements are available electronically in defined fields.

In other words, indicate whether data elements that are needed to compute the performance measure score are in defined, computer-readable fields.

[Response Begins]

Patient/family reported information (may be electronic or paper)

[Response Ends]

3.03. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using data elements not from electronic sources.

[Response Begins]

The Serious Illness Survey for Home-Based Programs assesses patient care experiences. During the field test of the survey, patients (or their family members, serving as proxy respondents) completed the survey by mail or telephone; web-based administration is possible for programs that collect patients' email addresses.

[Response Ends]

3.04. Describe any efforts to develop an eCQM.

[Response Begins]

N/A

[Response Ends]

3.06. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

[Response Begins]

The field test of the Serious Illness Survey for Home-Based Programs established the feasibility of identifying survey-eligible patients using programs' administrative records and using a survey vendor to conduct survey

sampling and data collection. As noted above in Section 2b.10, the overall survey response rate of 36.4% (30.4% in mail-only mode and 42.5% in mail with telephone follow-up) is comparable to, or better than, response rates observed for other care experience surveys in national use, suggesting that a survey of seriously ill patients is as feasible as surveys of other patient and caregiver populations.

[Response Ends]

Consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

3.07. Detail any fees, licensing, or other requirements to use any aspect of the measure as specified (e.g., value/code set, risk model, programming code, algorithm),

Attach the fee schedule here, if applicable.

[Response Begins]

The Serious Illness Survey for Home-Based Programs is made available under a Creative Commons license at no cost. All Serious Illness Survey materials are copyright RAND Corporation, 2021, and are made available under a Creative Commons Attribution—NonCommercial-NoDerivatives 4.0 International License. A complete set of survey resources, including guidance on administration, sampling, and analysis, and survey cover letters, mail questionnaires and telephone scripts in English and Spanish, is available at: at www.rand.org/Serious-Illness-Survey

[Response Ends]

4. Usability and Use

Extent to which potential audiences (e.g., consumers, purchasers, providers, policy makers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations.

Extent to which intended audiences (e.g., consumers, purchasers, providers, policy makers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement, in addition to demonstrating performance improvement.

4a.01. Check all current uses. For each current use checked, please provide:

- **Name of program and sponsor**
- **URL**
- **Purpose**
- **Geographic area and number and percentage of accountable entities and patients included**
- **Level of measurement and setting**

[Response Begins]

Not in use

[Not in use Please Explain]

Newly developed measure

[Response Ends]

4a.02. Check all planned uses.

[Response Begins]

Quality Improvement (internal to the specific organization)

[Response Ends]

4a.03. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing), explain why the measure is not in use.

For example, do policies or actions of the developer/steward or accountable entities restrict access to performance results or block implementation?

[Response Begins]

The measures are not currently in broad use because they are newly developed. There are no restrictions preventing administration of the survey. Survey materials are available at **www.rand.org/Serious-Illness-Survey**.

[Response Ends]

4a.04. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes: used in any accountability application within 3 years, and publicly reported within 6 years of initial endorsement.

A credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.

[Response Begins]

The Serious Illness Survey for Home-Based Programs was developed to address a measurement gap identified by serious illness stakeholders; these stakeholders recognized rapid expansion in the number of home-based serious illness care programs, and the potential introduction of alternative payment models that address the needs of the seriously ill. The measure developers are in ongoing discussions with home-based serious illness programs that are interested in implementing the survey to inform their internal quality improvement efforts; with payers that may incorporate measures from the survey in future initiatives that target the needs of serious illness populations; and with stakeholder organizations that are interested in hosting a database that would allow serious illness programs to voluntarily submit their survey response data to allow for benchmarking. The anticipated timeframe for introduction of new payer initiatives targeting serious illness populations is within 3 years; per the CMS Innovation Center's Strategy Refresh (CMS, 2021), "all models will consider or include patient-reported outcomes as part of the performance measurement strategy."

As of fall 2022, at least one large home-based serious illness program in North Carolina is administering the survey to allow for quality monitoring and inform quality improvement. In 2023, RAND will administer an adapted version of the Serious Illness Survey for Home-Based Programs to select Medicare Advantage members of Blue Cross and Blue Shield of Massachusetts to inform quality improvement and allow for comparison of care experiences between seriously ill individuals who do and do not receive specialized interventions.

Citation:

Centers for Medicare & Medicaid Innovation Center. (2021). Strategy Refresh. Available at: https://innovation.cms.gov/strategic-direction-whitepaper.

[Response Ends]

4a.05. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

Detail how many and which types of measured entities and/or others were included. If only a sample of measured entities were included, describe the full population and how the sample was selected.

[Response Begins]

The field test of the Serious Illness Survey for Home-Based Programs was conducted from October 2019 through January 2020 among patients of 32 geographically diverse serious illness programs that provide home-based care across the United States. As described in Section 2a.05, programs were recruited from a master list of 319 serious illness programs developed by the project team. To be eligible to participate, all programs needed to provide medical care to seriously ill patients in their homes. Almost all included programs provide after-hours access to care either by phone or in person and have either a physician or a nurse practitioner on the team that makes home visits. As described in Table 2a.1, programs were diverse with regard to geographic location in the United States, size (i.e., number of patients in care), and ownership. At the completion of the field test, programs were provided a summary report describing their survey results in comparison to all programs participating in the field test.

[Response Ends]

4a.06. Describe the process for providing measure results, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

[Response Begins]

At the conclusion of the Serious Illness Survey for Home-Based Programs field test, the project team provided a summary report via secure file transfer to each of the 32 participating programs. Reports included a description of the survey questionnaire, sampling, and administration procedures, as well as a table displaying the frequency of responses to each survey item for the given program, and for all programs participating in the field test. Inquiries about the field test, survey, and summary reports were addressed by the project team via email. The project team presented field test results to a technical expert panel composed of serious illness home-based program leaders, stakeholder groups, and quality measurement experts in June 2020.

[Response Ends]

4a.07. Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.

[Response Begins]

Programs participating in the field test were encouraged to provide feedback on their experience with the field test, the proposed final survey instrument, and results presented in their summary reports, and have reported that the survey will be useful for quality improvement activities. During a June 2020 virtual meeting, the project team sought feedback from the project's technical expert panel on survey items to include in the final version of the survey instrument and its quality measures. In addition, once the survey instrument was finalized and posted online, email blasts and press releases were distributed to inform potential users of the availability of the survey, and to seek feedback on survey content and administration.

[Response Ends]

4a.08. Summarize the feedback obtained from those being measured.

[Response Begins]

Programs participating in the field test reported that the data provided in the summary reports were useful and noted that they found requirements for participation in the field test (e.g., provision of sample data files) manageable. Other serious illness programs have reached out to the project team and expressed interest in fielding the survey, inquiring about whether any survey vendors are currently administering the survey, and the projected costs of administration.

[Response Ends]

4a.09. Summarize the feedback obtained from other users.

[Response Begins]

Members of the project's technical expert panel reviewed results from the field test, including the average and range of top-box scores on each evaluative survey item on the field test version of the survey, and advised the project team on items that would be important to retain in the final survey and those that should be removed. Panelists identified items that measured very similar concepts (e.g., an item on "talking as much as you wanted" about important health care decisions and another on "ever talking about what is important in your life"), and recommended that only a subset be selected for the final survey, and identified items that were either highly actionable for quality improvement (e.g., an item regarding involving family members or friends in discussions about health care as much as desired), or not considered actionable (e.g., an item regarding involving the patient in discussions about health care as much as desired).

[Response Ends]

4a.10. Describe how the feedback described has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

[Response Begins]

The project team refined the survey items included in the final survey instrument in response to feedback from the technical expert panel. For example, an item in the Communication domain regarding feeling "heard and understood" was maintained in the final measure to promote harmonization with an outpatient palliative care measure; an item in the Help for Symptoms domain regarding help for nausea and vomiting was dropped, as experts reported that this symptom was more likely to be addressed by providers (e.g., oncologists) other than the home-based program. Panelists noted the importance of ensuring that survey items could be used to assess telemedicine visits, which increased tremendously during the pandemic, in addition to in-person visits. The project team conducted cognitive interviews with six patients and family members from a diverse set of serious illness programs prior to finalizing the instrument. These cognitive interviews confirmed that respondents include both telemedicine and in-person visits in their evaluations of care.

[Response Ends]

4b.01. You may refer to data provided in Importance to Measure and Report: Gap in Care/Disparities, but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included). If no improvement was demonstrated, provide an explanation. If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

[Response Begins]

Serious Illness Survey for Home-Based Programs measures are newly developed and not currently in use. Programs and payers have expressed enthusiasm about the availability of the survey and its quality measures, noting that they might find it useful to administer the survey to assess quality of care in ongoing or future initiatives, and to inform quality improvement.

[Response Ends]

4b.02. Explain any unexpected findings (positive or negative) during implementation of this measure, including unintended impacts on patients.

[Response Begins]

We have not observed any unexpected findings.

[Response Ends]

4b.03. Explain any unexpected benefits realized from implementation of this measure.

[Response Begins]

As this is a newly developed measure that is not in broad use, we have not yet observed unexpected benefits from its implementation. However, during interviews conducted in the course of developing the field test survey, patients reported that it would be useful to be able to provide feedback on their care.

[Response Ends]

5. Comparison to Related or Competing Measures

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.

If you are updating a maintenance measure submission for the first time in MIMS, please note that the previous related and competing data appearing in question 5.03 may need to be entered in to 5.01 and 5.02, if the measures are NQF endorsed. Please review and update questions 5.01, 5.02, and 5.03 accordingly.

5.01. Search and select all NQF-endorsed related measures (conceptually, either same measure focus or target population).

(Can search and select measures.)

[Response Begins]

3665: Ambulatory Palliative Care Patients' Experience of Feeling Heard and Understood

3666: Ambulatory Palliative Care Patients' Experience of Receiving Desired Help for Pain

2651: CAHPS® Hospice Survey, Version 9.0

[Response Ends]

5.02. Search and select all NQF-endorsed competing measures (conceptually, the measures have both the same measure focus or target population).

(Can search and select measures.)

[Response Begins]

[Response Ends]

5.03. If there are related or competing measures to this measure, but they are not NQF-endorsed, please indicate the measure title and steward.

[Response Begins]

N/A

[Response Ends]

5.04. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s), indicate whether the measure specifications are harmonized to the extent possible.

[Response Begins]

Yes

[Response Ends]

5.05. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

[Response Begins]

N/A

[Response Ends]

5.06. Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality). Alternatively, justify endorsing an additional measure.

Provide analyses when possible.

[Response Begins]

2651 CAHPS Hospice Survey

The CAHPS Hospice Survey assesses care experiences of patients who died while receiving hospice care; respondents are informal caregivers of these patients. In contrast, respondents to the Serious Illness Survey for Home-Based Programs are living, seriously ill patients who are receiving care from a home-based serious illness program.

3665 Ambulatory Palliative Care Patients' Experience of Feeling Heard and Understood

3666 Ambulatory Palliative Care Patients' Experience of Receiving Desired Help for Pain

Measures 3665 and 3666, developed by the American Academy of Hospice and Palliative Medicine, assess communication and pain palliation among patients who have received specialty palliative care in an outpatient setting. In contrast, respondents to the Serious Illness Survey for Home-Based Programs are patients who have received care from a special program that provides care to seriously ill patients in their homes. The Serious Illness Survey for Home-Based Programs Communication measure includes a survey item regarding feeling "heard and understood;" this harmonizes with measure 3665, Ambulatory Palliative Care Patients' Experience of Feeling Heard and Understood, but also includes survey items regarding other aspects of communication. Similarly, the Serious Illness Survey for Home-Based Programs Help for Symptoms measure includes a survey item regarding getting desired help for pain; this harmonizes with measure 3666, Ambulatory Palliative Care Patients' Experience of Receiving Desired Help for Pain, but also includes other survey items regarding desired help for trouble breathing and anxiety or sadness.

[Response Ends]

Appendix

Supplemental materials may be provided in an appendix.:

Available in attached file

Attachment: 3726_3726_Table 1b.02b Serious Illness Survey for Home-Based Program Measures Deciles 2022_11-508.xlsx

Contact Information

Measure Steward (Intellectual Property Owner): RAND Corporation

Measure Steward Point of Contact: Anhang Price, Rebecca, ranhangp@rand.org

Measure Developer if different from Measure Steward: RAND Corporation

Measure Developer Point(s) of Contact: Anhang Price, Rebecca, ranhangp@rand.org

Additional Information

1. Provide any supplemental materials, if needed, as an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be collated one file with a table of contents or bookmarks. If material pertains to a specific criterion, that should be indicated.

[Response Begins]

Available in attached file

[Response Ends]

Attachment: 3726_3726_Table 1b.02b Serious Illness Survey for Home-Based Program Measures Deciles 2022_11-508.xlsx

2. List the workgroup/panel members' names and organizations.

Describe the members' role in measure development.

[Response Begins]

The Technical Expert Panel for development of the Serious Illness Survey for Home-Based Programs met in 2018 to discuss the design of the field test and priority topics for inclusion in the draft survey instrument, and in 2020, to review the results of the field test and provide feedback on final survey content. The panel included the following individuals. Affiliations noted were current as of the June 2020 meeting.

- Paul Cleary, PhD, Panel Chair, Yale School of Public Health
- Katherine Ast, MSW, LCSW, American Academy of Hospice and Palliative Medicine
- Lori Bishop, MHA, BSN, RN, CHPN, National Hospice and Palliative Care Organization
- Nikki Davis, MSN, FNP-C, GNP-BC, ACHPN, Aspire Health
- Susan Edgman-Levitan, PA-C, Massachusetts General Hospital
- Betty Ferrell, RN, PhD, MA, FAAN, FPCN, CHPN, City of Hope Medical Center
- Torrie Fields, MPH, Votive Health
- Mark Friedberg, MD, MPP, Blue Cross Blue Shield of Massachusetts
- Amy Kelley, MD, MSHS, Icahn School of Medicine, Mount Sinai
- Rebecca A. Kirch, JD, National Patient Advocate Foundation
- Dana Lustbader, MD, FAAHPM, ProHEALTH, Optum Supportive Care
- Diane Meier**,** MD, FACP, Center to Advance Palliative Care
- Sarah Scholle, DrPH, National Committee for Quality Assurance
- Bradley Hall Stuart, MD, MMM, FAAHPM, Coalition to Transform Advanced Care
- Joan Teno, MD, MS, Project Team Advisor, OHSU School of Medicine
- Rodney O. Tucker, MD, University of Alabama at Birmingham

[Response Ends]

3. Indicate the year the measure was first released.

[Response Begins]

2021

[Response Ends]

4. Indicate the month and year of the most recent revision.

[Response Begins]

N/A (not revised)

[Response Ends]

5. Indicate the frequency of review, or an update schedule, for this measure.

[Response Begins]

There is no set update schedule for this measure.

[Response Ends]

6. Indicate the next scheduled update or review of this measure.

[Response Begins]

There is no set update schedule for this measure.

[Response Ends]

7. Provide a copyright statement, if applicable. Otherwise, indicate "N/A".

[Response Begins]

[Response Ends]

8. State any disclaimers, if applicable. Otherwise, indicate "N/A".

[Response Begins]

N/A

[Response Ends]

9. Provide any additional information or comments, if applicable. Otherwise, indicate "N/A".

[Response Begins]

N/A

[Response Ends]