
CBE ID

1623

Title

Bereaved Family Survey - Performance Measure (BFS-PM) Score (%) for all Veteran Affairs
Medical Center Inpatient Deaths

Project

Advanced Illness and Post-Acute Care

Endorsement Status

Endorsed

Is Under Review

No

Next Maintenance Cycle

Fall 2029

Previous Endorsement Cycle

Fall 2024

Initial Endorsement

Tue, 02/14/2012 - 09:45

Steward

Department of Veterans Affairs / Hospice and Palliative Care

1.0 New or Maintenance

Maintenance

1.1 Measure Structure

Single Measure

1.3 Electronic Clinical Quality Measure (eCQM)

No

1.6 Measure Description

The Bereaved Family Survey-Performance Measure (BFS-PM) is an outcome measure that is used to assess overall quality of care in the last month of life. Currently, the BFS is administered to the next-of-kin of all Veterans who die in a VA inpatient setting (i.e., acute units, intensive care units, inpatient hospice and palliative care units, and VA nursing homes) 4-6 weeks post-death. The BFS-PM is calculated using the global rating item included on the 20-item BFS that has separate versions for male and female Veterans and is available in English and Spanish. The BFS global rating item asks: "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 the care [he/she] received in the last month

of life?” The BFS-PM is calculated as the proportion of family members who provided a “top box” rating of 9 or 10 vs. 0-8 on the global rating item. BFS-PM scores are used for the purposes of monitoring quality of care for Veterans at the end of life nationally, facility benchmarking within the VA healthcare system, and targeting quality improvement efforts.

1.7 Composite Measure

No

1.7 Measure Type

Patient-reported Outcome Performance Measure (PRO-PM)

1.8 Level of Analysis

Facility

1.8b Other Level of Analysis

Individual Patient

1.9 Care Setting

Other

1.9b Other Care Setting

VA inpatient facilities (includes acute units, intensive care units, inpatient hospice and palliative care units, and VA nursing homes)

1.10 Measure Rationale

Multiple National Academies (formerly the Institute of Medicine) consensus reports have underscored the degree to which the quality of end-of-life (EOL) care in the United States needs to be improved.^{1,2} The challenges of EOL care are particularly significant in the U.S. Department of Veterans Affairs (VA) healthcare system because the VA provides care for an increasingly older population with multiple comorbid conditions. Approximately 200,000 Veterans in the U.S. die each year and this number is expected to increase.³ Currently, almost half (49%) of the population of 9 million VA-enrolled Veterans are over the age of 65, and 30% of these Veterans are over the age of 80.⁴ These demographic trends mean that, like other healthcare systems, there is a great need for high-quality EOL care that is patient- and family-centered.

For over a decade, the Bereaved Family Survey-Performance Measure (BFS-PM) has been VA’s primary measure of the quality of EOL care provided to Veterans. The BFS-PM is implemented system-wide which allows for the comparison of the quality of EOL care delivered across all VA inpatient facilities nationwide and the identification of opportunities for improvement. The BFS-PM is also used to monitor the effectiveness of quality improvement efforts to improve EOL care locally and nationally. Finally, the BFS-PM is used by VA to recognize facilities that provide outstanding EOL care, so that successful processes and structures of care can be identified and disseminated.

The BFS-PM is calculated using the global rating item included on the 20-item BFS that has separate versions for male and female Veterans and is available in English and Spanish. Currently, the BFS is administered to the next-of-kin of all Veterans who die in a VA inpatient setting (i.e., acute units, intensive care units (ICUs), inpatient hospice units, and VA Community Living Centers (CLCs; VA nursing homes)) 4-6 weeks post death. Post-death surveys of family members, such as the BFS, are an essential strategy for assessing the quality of EOL care. There are at least 5 reasons why post-death family surveys are an essential part of an effective measurement strategy. First, family surveys can assess the care of all Veterans who die, even those whose prognosis is uncertain, and who therefore might not be identified as “terminally ill” in a prospective assessment. Second, family surveys also avoid challenges of data collection from patients near EOL, in whom cognitive impairment is common, particularly in ICU deaths. Third, family surveys can retrospectively assess care within a few weeks of death instead of asking at the time of death – an emotionally difficult time when data collection from patients or families may be felt to be unacceptably intrusive. Fourth, families’ assessments offer an essential source of data to assess the support that is provided to family members themselves. For these reasons, post-death assessments by family members offer an essential source of data that define the quality of EOL care that VA can provide for Veterans and their family members.

References

1. Committee on Approaching Death: Addressing Key End of Life Issues; Institute of Medicine. Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life. Washington (DC): National Academies Press (US); 2015
2. National Academies of Sciences, Engineering, and Medicine. 2022. The national imperative to improve nursing home quality: Honoring our commitment to residents, families, and staff. Washington, DC: The National Academies Press.
3. Agency for Healthcare Research and Quality. MEPSnet Query: Annual mortality estimates among older users of VHA Users, 2018-2022. Accessed 4/16/2024.
4. U.S. Veteran Service Support Center (VSSC). Enrollment pyramid [internal VA website]. Accessed October 2024.

1.13 Data Dictionary

Not attached. I attest that all information will be provided where codes and/or value sets are needed (1.14a - 1.15c).

1.13a Attach Data Dictionary

[1.13. Data Dictionary.xlsx](#)

1.14 Numerator

The numerator includes the number of respondents who gave a “top box” score of 9 or 10 on the BFS global rating item.

1.14a Numerator Details

The BFS global item asks " 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 the care [he/she] received in the last month of life?

1.15 Denominator

The denominator includes the total number of completed surveys among Veterans who died as an inpatient in a VA facility (defined as surveys with a valid response for the BFS global rating item plus at least 12 more valid responses on the forced-choice items).

1.15a Denominator Details

Surveys are administered 4-6 weeks after the Veteran’s death and are received on a rolling basis.

1.15b Denominator Exclusions

Bereaved family members (i.e., next-of-kin (NOK)) of Veterans who: 1) died within 24 hours of admission (unless the Veteran had a previous hospitalization of at least 24 hours in the last month of life), and 2) did not die in an inpatient VA facility are excluded.

Bereaved family members (i.e., NOK) may also be excluded if: 1) they (or a proxy) do not feel informed enough to answer questions about the Veteran's last month of life (determined by family member’s report); 2) they cannot be contacted; 3) they decline to participate in the survey or do not agree to be interviewed; 4) they feel too uncomfortable to answer the survey questions, or they are reluctant to discuss details about the Veteran's death; 5) they do not speak English or Spanish; and 6) did not have a medical record with complete NOK contact information, including a working telephone number and complete address).

1.15c Denominator Exclusions Details

Name, address, and phone number of the Veteran’s next-of-kin (NOK) or emergency contact are

required for determining denominator exclusions, in addition to information regarding the Veteran's terminal admission and any prior admissions during the last 31 days of life, including length of stay. This information is obtained from VA's Corporate Data Warehouse (CDW) which stores Veterans' administrative demographic and clinical data.

1.15d Age Group

Other

1.15e Age Range in Years

Adults aged 18-103 years

1.16 Type of Score

Rate/proportion

1.17 Measure Score Interpretation

Better performance = Higher score

1.18 Calculation of Measure Score

The BFS-PM is calculated using the BFS global rating item which asks: "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 the care [he/she] received in the last month of life?"

- First, responses on the BFS global rating are re-coded as either "1" (rating of 9 or 10) or "0" (rating of 0-8). Thus, a value of "1" indicates that the family member reported that they and/or the Veteran received the best care possible (i.e., a "top box" rating). A value of "0" reflects all other possible responses (i.e., rating of 0-8). Items are coded as missing if respondents cannot or refuse to answer the item.
- The BFS-PM is calculated at the facility-level by dividing the total number of BFS respondents who reported a "top box" rating of 9 or 10 on the BFS global rating item (numerator) by the total number of completed surveys among Veterans who died as an inpatient in the facility (denominator).
- Completed surveys are defined as those with a valid response for the BFS global rating item plus at least 12 more valid responses on the forced-choice items.
- This scoring system produces facility-level proportions that when multiplied by 100 reflects the percentage of Veterans who received the best possible overall care (i.e., BFS-PM).
- BFS-PM scores can also be calculated nationally.
- BFS-PM scores are weighted for nonresponse and patient case mix prior to reporting.
- Scores are reported quarterly using the rolling average of the previous four quarters of data.

1.19 Measure Stratification Details

The BFS-PM is reported as a facility-level measure that combines all venues of death: acute care units, intensive care units, inpatient hospice units, and Community Living Center (CLC; VA nursing home). The BFS-PM is also stratified by venue of death for quality improvement purposes.

The stratification variable “hcldeath” (inpatient hospice/CLC death vs. acute/ICU death) is highlighted and described in the Data Dictionary (see section 1.13). See section 4.4.5 and 4.4.5a for Discrimination and Calibration results of the risk-adjustment models for the BFS-PM stratified by venue of death.

1.20 Types of Data Sources

Administrative Data, Patient-Reported Data and/or Survey Data

1.21a Data Collection Tool URL(s)

https://www.cherp.research.va.gov/Veteran_Experience_Center/Bereaved_Family_Sur...

1.21b Attach Data Collection Tool(s)

[1.21b. Data Collection Tools.zip](#)

1.22 Proxy Responses

Yes

1.23 Survey Respondent

Family or Other Caregiver

1.24 Data Collection and Response Rate

For each Veteran, we identify one contact (generally a family member) in the following order of priority: 1) individual named as the patient's next-of-kin (NOK) in VA's Central Patient Record System; 2) individual named as the patient's secondary NOK; 3) individual named as durable power of attorney for health care (DPOA) or emergency contact.

Four to six weeks after the Veteran's death, family members are mailed a paper copy of the Bereaved Family Survey (BFS) and a letter from the director of the Veteran Experience Center, Dr. Cari Levy, describing the survey. Family members are asked to return the paper copy of the survey, complete the survey online, or call the Veteran Experience Center to complete the BFS over the phone with a staff member. If we do not receive a completed survey by mail, we send out a second survey in the mail with a reminder letter. If we have still not received a completed survey by mail, we send out a reminder postcard with our contact information (for completing the survey over the phone, or to receive another mailed copy of the survey) and instructions for completing the survey online. Family members who have not responded to the survey after all three mail contacts receive a reminder follow-up phone call.

If a family member receives the BFS and they feel that they have been contacted in error, or if they feel for any reason that they are not equipped to fill out the BFS, they are asked to contact the Veteran Experience Center by phone to help them understand why they were contacted, and for assistance in potentially identifying another family member or friend who could complete the

BFS instead.

Response rates are calculated by removing all deaths that occurred within 24 hours of admission, then dividing the number of completed surveys by the total number of eligible NOK of Veterans who died as an inpatient and who were admitted for at least 24 hours in the last month of life (exclusion criteria are provided in section 1.15b). For reporting, a table providing a breakdown of survey responses/nonresponse (e.g. completed surveys, ineligible deaths) is provided with scores along with instructions on how to calculate the response rate.

Nationally in FY22-23 (October 1, 2021- September 30, 2023), the breakdown of completed and noncompleted surveys (including reasons for non-response, if provided), and reasons for BFS ineligibility were as follows: completed survey=11,568 (39.8%), did not return survey/answer phone=10,948 (37.6%), declined survey=524 (1.8%), did not know enough about care=307 (1.1%), reluctant to talk about death=180 (0.6%), incomplete survey=143 (0.5%), Veteran had a <24 hour admission=1,194 (4.1%), incorrect/missing contact information=4,077 (14.0%), and no NOK listed/did not speak English/Spanish=146 (0.5%).

1.25 Data Source Details

In addition to the Bereaved Family Survey (BFS), the VA Clinical Data Warehouse is used to calculate the case-mix and nonresponse adjustments for the BFS-Performance Measure (BFS-PM). Variables are extracted directly from the VA Corporate Data Warehouse (an integrated system of national VA databases including clinical, electronic medical record, administrative and financial data) using standardized algorithms. The CDW is updated daily and has been validated and used extensively for quality improvement and research.¹

References

1. Price LE, Shea K, Gephart S. The Veterans Affairs Corporate Data Warehouse: Uses and Implications for Nursing Research and Practice. Nurs Adm Q. 2015 Oct-Dec;39(4):311-8.

1.26 Minimum Sample Size

A minimum sample size of 30 respondents is suggested to make comparisons between groups (i.e., facilities). For example, in FY22 (October 1, 2021- September 30, 2022), 81 out of 145 (55%) facilities had >29 respondents over the four-quarter reporting period. We are careful to advise caution to these facilities when interpreting results. Our internal facility-level reports include a clear indicator (*) for facilities with less than 30 responses, noting that the sample size is too small for confident interpretation.

2.1 Attach Logic Model

[2.1 Logic Model.docx](#)

2.2 Evidence of Measure Importance

Care provided to patients at the end of life (EOL) should be patient-centered and family-oriented.¹ Thus, a common quality indicator for EOL care is the patient and family experience of care. The Bereaved Family Survey – Performance Measure (BFS-PM) represents one of few existing measures of organizational performance in EOL care quality that has been collected consistently to monitor quality and to identify facilities that require improvement. Since 2010, the BFS-PM has been used to evaluate the quality of care in all 145 VA inpatient facilities nationally, including 122 VA nursing homes (also known as Community Living Centers). Notably, the recent National Academy of Medicine³ report on improving nursing home quality in the United States highlighted the BFS as one of the largest system-level efforts to measure family evaluations of the quality of EOL care provided in nursing homes. In evaluations spanning over a decade, the BFS-PM (and the BFS overall rating item from which the measure is derived) has been linked to several care structures and processes. See **section 7.1, Supplemental Attachment, Appendix A** for a full list of BFS-related publications).

The BFS-PM has been used as a key outcome metric in evaluations of EOL care structures and facility characteristics. In a study by Ersek and colleagues,² BFS overall rating scores collected from family members of Veterans who died in VA nursing homes (i.e. Community Living Centers) were higher than those who died in acute or intensive care units. In another evaluation, Rolnick and colleagues⁴ found that bereaved family members of Veterans who died in intensive care units rated the quality of EOL care significantly higher than families of Veterans who died on general acute care units. The BFS overall rating has also been linked to nurse staffing levels in acute care units in VA Medical Centers.⁵

Several care processes have also been linked to the BFS-PM (see section 4.3.3-4.3.5, Validity) and the BFS overall rating. One set of these care processes, obtained from Veteran medical records, have been identified as best practices in the National Consensus Project's Clinical Practice Guidelines for Quality Palliative Care.⁶ These processes include: 1) receipt of a comprehensive palliative care consultation in the patient's final months of life; 2) death in a dedicated inpatient hospice unit; 3) patient/family contact with a chaplain in the last month of life; and 4) evidence of emotional support given to a family member up to one month post-Veteran death via a bereavement contact. Receipt of palliative care is associated with higher BFS overall ratings, with the highest ratings observed among Veterans who receive a consultation earlier in their illness trajectory.⁷⁻⁹ Similarly, receipt of hospice care is associated with higher BFS overall ratings.¹⁰⁻¹¹

In addition to chart-derived process measures, the BFS includes three validated factors, or subscales, that also measure EOL care processes and include: Respectful Care and

Communication (e.g., staff attended to personal care needs), emotional and spiritual support (e.g., staff gave enough emotional support before death), receipt of death benefit information (e.g., staff gave enough information about survivor benefits).⁹ Bereaved family members who report higher scores on each of these factors are more likely to report an optimal, or “top box” response on the BFS overall rating.⁹

References

1. National Academies of Sciences, Engineering, and Medicine. 2015. Dying in America: Improving Quality and Honoring Individual Preferences Near the End of Life. Washington, DC: The National Academies Press. <https://doi.org/10.17226/18748>.
2. National Academies of Sciences, Engineering, and Medicine 2022. The National Imperative to Improve Nursing Home Quality: Honoring Our Commitment to Residents, Families, and Staff. Washington, DC: The National Academies Press. <https://doi.org/10.17226/26526>.
3. Ersek M, Thorpe J, Kim H, Thomasson A, Smith D. Exploring End-of-Life Care in Veterans Affairs Community Living Centers. J Am Geriatr Soc. 2015 Apr;63(4):644-50. doi: 10.1111/jgs.13348.
4. Rolnick JA, Ersek M, Wachterman MW, Halpern SD. The Quality of End-of-Life Care among ICU versus Ward Decedents. Am J Respir Crit Care Med. 2020 Apr 1;201(7):832-839.
5. Kutney-Lee A, Brennan CW, Meterko M, Ersek M. Organization of nursing and quality of care for Veterans at the end of life. J Pain Symptom Manage. 2015;49(3):570-7.
6. National Consensus Project for Quality Palliative Care. Clinical Practice Guidelines for Quality Palliative Care, 4th edition. Richmond, VA: National Coalition for Hospice and Palliative Care; 2018. <https://www.nationalcoalitionhpc.org/ncp>.
7. Carpenter JG, McDarby M, Smith D, Johnson M, Thorpe J, Ersek M. Associations between Timing of Palliative Care Consults and Family Evaluation of Care for Veterans Who Die in a Hospice/Palliative Care Unit. J Palliat Med. 2017 Jul;20(7):745-751. doi: 10.1089/jpm.2016.0477

8. Casarett D, Pickard A, Bailey FA, et al. Do palliative consultations improve patient outcomes? J Am Geriatr Soc. Apr 2008;56(4):593-599.

9. Thorpe JM, Smith D, Kuzla N, Scott L, Ersek M. Does Mode of Survey Administration Matter? Using Measurement Invariance to Validate the Mail and Telephone Versions of the Bereaved Family Survey. J Pain Symptom Manage. 2016 Mar;51(3):546-56. doi: 10.1016/j.jpainsymman.2015.11.006.

10. Feder S, Ersek M, Kutney-Lee A, Luhrs C, Bastian L, Akgun K. End-of-Life Care Quality Varies Among Facilities in the Department of Veterans Affairs for Patients with Heart Failure. J Pain Symptom Manage. 2020 59(2):546-7. doi: <https://doi.org/10.1016/j.jpainsymman.2019.12.294>.

11. Richards CA, Hebert PL, Liu CF, Ersek M, Wachterman MW, Taylor LL, Reinke LF, O'Hare AM. Association of Family Ratings of Quality of End-of-Life Care With Stopping Dialysis Treatment and Receipt of Hospice Services. JAMA Netw Open. 2019 Oct 2;2(10):e1913115. doi: 10.1001/jamanetworkopen.2019.13115.

2.4 Performance Gap

Bereaved Family Survey data from FY22 (October 1, 2021-September 30, 2022) were used for the performance gap analysis. Scores were adjusted using weighting to account for differences in case-mix and nonresponse bias. These weights were calculated using additional data obtained for the same time period from VA's Corporate Data Warehouse (CDW) which stores Veterans' administrative demographic and clinical data. The performance gap analysis included 145 measured entities (facilities) and 5,812 BFS responses.

Table 1. Performance Scores by Decile

	Performance Gap												
	Overall	Minimum	Decile_1	Decile_2	Decile_3	Decile_4	Decile_5	Decile_6	Decile_7	Decile_8	Decile_9	Decile_10	Maximum
Mean Performance Score	75.2%	28%	52.3%	62.1%	65.7%	70.7%	74.5%	77.7%	81.1%	84.8%	88.1%	96.4%	100.0%
N of Entities	145	1	15	14	15	14	15	14	15	14	15	14	7
N of Persons													
/ Encounters	5,812	5	236	451	502	505	557	642	684	709	745	781	35
/ Episodes													

2.6 Meaningfulness to Target Population

Many BFS respondents often express their appreciation for being asked for their feedback on the quality of care that they and their Veteran family member received via the BFS. The consistently

high response rate (average 44% over 10 years) to the BFS is also an expression of the value placed on the survey by bereaved family members.

3.1 Contributions Towards Closing Care Gaps

While differences in BFS-PM scores across VA facilities by socio-contextual variables have not been tested formally, the BFS overall rating (which informs the BFS-PM) has been used extensively at the individual-level to explore such differences and contribute to advancing health equity.

Using BFS data collected over a five-year period (October 2009-September 2014), Kutney-Lee and colleagues¹ documented sizable differences in bereaved family perceptions of the quality of end-of-life (EOL) care by Veteran race and ethnicity. In a study of over 51,000 Veterans, the team found that families of Black and Hispanic Veterans were significantly less likely than those of non-Hispanic White Veterans to provide a “top box” score on the BFS overall rating item, and this difference was largest for families of non-Hispanic Black Veterans (48% vs. 62%, $p<0.001$). Follow-up studies have been conducted to explore these differences and inform the development of quality improvement initiatives.^{2,3} The VA Hospice and Palliative Care Program Office plans to explore these differences at the facility-level and may consider the creation of an Annual Disparities Report that would provide BFS-PM ratings stratified by race and ethnicity.

Differences in BFS overall ratings have also been examined by sex, rurality, and war era (which is also a reasonable proxy for age). In a study by Ersek and colleagues,⁴ bereaved family members of women Veterans were more likely to provide a “top box” rating on the BFS overall rating item (63% vs. 56%, $p=0.003$). In a study examining the role of urbanity/rurality on quality of EOL care, Del Rosario and colleagues⁵ found no differences in the BFS overall ratings between family members of Veterans who lived in rural and urban areas. Due to this work, VA is exploring more community resources and clinical resource hubs to improve access to care for rural Veterans as these Veterans were less likely to die in an inpatient hospice unit. Finally, Kutney-Lee et al.⁶ compared quality of EOL care ratings for Vietnam and World War II/Korean War-era Veterans and found no statistically significant differences between groups in the percentage of bereaved family members who provided a “top box” BFS overall rating (62% vs. 62%, $p=0.276$). However, this work raised awareness of the increased mental health needs among Vietnam-era Veterans, particularly near EOL.

References

-

1. Kutney-Lee A, Smith D, Thorpe J, Del Rosario C, Ibrahim S, Ersek M. Race/Ethnicity and End-of-Life Care Among Veterans. *Med Care*. 2017 Apr;55(4):342-351. doi:

10.1097/MLR.0000000000000637.

2. Kutney-Lee A, Bellamy SL, Ersek M, Medvedeva EL, Smith D, Thorpe JM, Brooks Carthon JM. Care processes and racial/ethnic differences in family reports of end-of-life care among Veterans: A mediation analysis. *J Am Geriatr Soc.* 2022 Apr;70(4):1095-1105. doi: 10.1111/jgs.17632.

3. Kutney-Lee A, Rodriguez KL, Ersek M, Carthon JMB. "They Did Not Know How to Talk to Us and It Seems That They Didn't Care:" Narratives from Bereaved Family Members of Black Veterans. *J Racial Ethn Health Disparities.* 2023 Sep 21. doi: [10.1007/s40615-023-01790-4](https://doi.org/10.1007/s40615-023-01790-4).

4. Ersek M, Smith D, Cannuscio C, Richardson DM, Moore D. A nationwide study comparing end-of-life care for men and women veterans. *J Palliat Med.* 2013 Jul;16(7):734-40. doi: 10.1089/jpm.2012.0537.

5. Del Rosario C, Kutney-Lee A, Sochalski J, Ersek M. Does Quality of End-of-Life Care Differ by Urban-Rural Location? A Comparison of Processes and Family Evaluations of Care in the VA. *J Rural Health.* 2019 Sep;35(4):528-539. doi: 10.1111/jrh.12351.

-

6. Kutney-Lee A, Smith D, Griffin H, Kinder D, Carpenter J, Thorpe J, Murray A, Shreve S, Ersek M. Quality of end-of-life care for Vietnam-era Veterans: Implications for practice and policy. *Healthc.* 2021 Jun;9(2):100494. doi: 10.1016/j.hjdsi.2020.100494.

4.1 Feasibility Assessment

The feasibility of identifying eligible family members to receive the BFS and collecting the survey is well-established. First, procedures have been developed to permit an efficient and accurate identification of deaths as well as risk-adjustment and nonresponse bias adjustment variables from VA's electronic health record in the Corporate Data Warehouse (CDW). These include a system of checks to ensure that Veterans are deceased, eligible, and that the data collected from CDW is valid and accurate. Second, we have refined contact procedures to maximize mailed survey/interviewer efficiency, thereby decreasing costs. Third, we have developed operating procedures for addressing unresolved issues that are identified during data collection /interviews. These processes allow staff/interviewers to make rapid referrals to the appropriate VA resources to provide assistance to bereaved family members (e.g. for assistance with burial or funeral benefits, bereavement support). For example, during the two-year time period of data analyzed for this submission, staff made 40 referrals for bereavement support, 68 referrals to follow-up

with concerns about the care received, 14 referrals for questions about VA benefits, and 24 referrals for other concerns.

BFS response rates and missing data: Response rates are calculated by removing all deaths that occurred within 24 hours of admission, then dividing the number of completed surveys by the total number of eligible NOK of Veterans who died as an inpatient and who were admitted for at least 24 hours in the last month of life (exclusion criteria are provided in section 1.15b). For reporting, a table providing a breakdown of survey responses/nonresponse (e.g. completed surveys, ineligible deaths) is provided with scores along with instructions on how to calculate the response rate. Nationally in FY22 and FY23 (October 1, 2021- September 30, 2023), the breakdown of completed and noncompleted surveys (including reasons for non-response, if provided), and reasons for BFS ineligibility were as follows: completed survey=11,568 (39.8%), did not return survey/answer phone=10,948 (37.6%), declined survey=524 (1.8%), did not know enough about care=307 (1.1%), reluctant to talk about death=180 (0.6%), incomplete survey=143 (0.5%), Veteran had a <24 hour admission=1,194 (4.1%), incorrect/missing contact information=4,077 (14.0%), and no NOK listed/did not speak English/Spanish=146 (0.5%).

4.3 Feasibility Informed Final Measure

The Bereaved Family Survey (BFS) has been in use for over a decade. Therefore, feasibility assessments are continuously ongoing. Over the lifespan of the BFS, changes in the availability of data have occurred as well as the emergence of new data sources. Additionally, new methods for obtaining more accurate and efficient data capture have been tested and adopted. Data cleaning procedures are in place and data are consistently checked both externally by the survey vendor as well as internally by the Veteran Experience Center team. The BFS has also increased performance reporting exposure over the past decade and is now included on several VA program Dashboards, publicly reported on the VA Geriatrics and Extended Care website and is an official VA Performance Measure included in the performance evaluations of VA leadership staff. Quarterly estimates of the cost of administering, analyzing and report the results of the BFS have been accurate since 2013. Prior to 2013, cost estimates were too high, and the survey mode was changed from a phone interview to a paper mail survey to reduce costs.

4.4 Proprietary Information

Not a proprietary measure and no proprietary components

5.1.1 Data Used for Testing

Bereaved Family Survey (BFS) data from FY22 and FY23 (October 1, 2021-September 30, 2023) were used for testing. Additional data was obtained for the same time period from VA's Corporate Data Warehouse (CDW) which stores Veterans' administrative demographic and clinical data.

Data to evaluate person-level test-retest reliability were collected between January 1, 2024 and March 31, 2024 by mail and phone from 116 randomly selected BFS-eligible next-of-kin who agreed to complete the BFS on a second occasion (30 days apart).

5.1.2 Differences in Data

Reliability Testing: Person-level, test-retest reliability testing used BFS data collected between January 1, 2024 and March 31, 2024 by mail and phone from 116 randomly selected BFS-eligible next-of-kin who agreed to complete the BFS on a second occasion (30 days apart).

Person-level data element reliability and accountable entity-level reliability testing used BFS data from FY22 (October 1, 2021-September 30, 2022). These tests were conducted using a single year of data (FY22) to reflect the same length of time used for BFS-PM score calculation (i.e., four quarters, or one fiscal year, of data).

Accountable entity-level test-retest reliability (year-to-year stability) testing used BFS and CDW data from FY22 and FY23 (October 1, 2021-September 30, 2023).

Validity Testing: All person- and facility-level validity testing used BFS and CDW data from FY22 and FY23 (October 1, 2021-September 30, 2023). For validity testing, we maximized our sample size by pooling data from both fiscal years.

Nonresponse bias was evaluated using CDW data from FY22 (October 1, 2021-September 30, 2022) for Veterans whose next-of-kin did not complete a BFS.

Risk Adjustment/Calibration and Discrimination Testing: Risk adjustment modeling and associated calibration/discrimination testing used BFS and CDW data from FY22 (October 1, 2021-September 30, 2022) to reflect the same length of time used for BFS-PM score calculation (i.e., four quarters, or one fiscal year, of data).

5.1.3 Characteristics of Measured Entities

Our sample included 146 facilities for facility-level testing. Facilities were distributed fairly equally across U.S. regions: East (26.0%), Midwest (24.0%), South (29.5%), and West (20.5%). Nearly two-thirds (64.4%) of facilities were classified as high complexity. Within these facilities, there were 11,568 Veterans with an associated BFS response in FY22 and FY23. The largest proportion of Veterans were cared for in facilities in the southern U.S (4,074; 35.2%) and were admitted to a high complexity facility (9,424; 81.5%). See **section 7.1, Supplemental Appendix B, Table 1 (p.1)** for full results.

5.1.4 Characteristics of Units of the Eligible Population

The characteristics of each of the testing samples are provided in **section 7.1, Supplemental Appendix B, Table 2 (p.1-3)**, including the test-retest reliability sample (n=116), and BFS nonresponders (n=16,325) and BFS responders (n=11,568) across FY22 and FY23.

Veterans whose next-of-kin who were included in the test-retest sample (n=116) were, on average, 80.5 years old at the time of death and the majority were male (97.5%) and non-Hispanic White (83.6%). The most common primary diagnosis category at the time of death was infectious/parasitic disease (19.8%).

Compared to Veterans that had a BFS response (n=11,568), Veterans whose next-of-kin did not respond, i.e. BFS nonresponders (n=16,325), were more likely to be younger at time of death (75.8 vs. 78.2) and to be of non-Hispanic Black race (21.3% vs. 15.7%) or Hispanic ethnicity (7.5% vs. 5.2%). The most common primary diagnosis category across BFS nonresponders and responders were infectious and parasitic diseases (22.9% and 19.0%, respectively) and neoplasms (20.3% and 20.8%, respectively).

5.2.1 Level(s) of Reliability Testing Conducted

Person or encounter level (i.e., data element) (e.g., inter-abstractor reliability), Accountable entity level (i.e., measure score) (e.g., signal-to-noise analysis)

5.2.2 Method(s) of Reliability Testing

1. Person-Level

-

Test-Retest Reliability Testing

Overview of Test-Retest Reliability Testing Sample: To evaluate test-retest reliability as a measure of repeatability (i.e., stability) of the BFS overall rating over time, we randomly selected 116 BFS respondents who agreed to complete the BFS on a second occasion (30 days apart).

Statistical Analysis #1: Test-Retest Reliability of Absolute Agreement (Cohen's kappa):

For binary outcomes such as the overall BFS score, Cohen's kappa is a commonly used measure of absolute agreement between scores measured by multiple raters, or scores measured by stability test-retest agreement within a single respondent measured over multiple occasions. Or alternatively, Cohen's kappa assesses the ability of an item to produce the same score on multiple occasions. Cohen's kappa measures the within-respondent error in test-retest repeatability of the overall BFS score.

Statistical Analysis #2: Test-Retest Reliability of Absolute Agreement (two-way random effects, absolute agreement, single rater/measurement): The intraclass correlation coefficient (ICC) is a desirable measure of test-retest reliability because it reflects both the degree of correlations and absolute agreement between time1 and time2 measurement occasions. While ICCs were developed for test-retesting of continuous outcomes, recent research suggests ICCs on binary outcomes reasonably approximates traditional, linear ICC estimates. We chose ICC measures of absolute agreement in addition to Cohen's kappa because the kappa coefficient may yield downward-biased reliability estimates depending the prevalence and variability of responses in the sample on the low values when the ratings suggest high reproducibility ("kappa paradox").¹ We examined ICC test-retest absolute agreement with a two-way random intercept ICC (2,1). ICC(2,1) assumes BFS respondents were randomly selected from the larger population of respondents with similar characteristic.

Statistical Analysis #3: Test-Retest Reliability (Logistic Regression): Strength of association between time 1 and time 2 BFS overall scores was evaluated using a logistic regression of the following form:

$$\log (Pr(BFS=1 | time=2)/1-(Pr(BFS=1 | time=2)))=B_0 + B_1(BFS_{time=1})$$

The exponentiation of provides an estimate of the odds ratio. The odds ratio may be interpreted as follows: Compared to those who reported BFS=0 at time 1, respondents who reported BFS=1 at time 1 had X times the odds of reporting BFS=1 at time 2.

Data Element Reliability Testing

Factor Analysis Approach to Evaluating Single-Item Reliability: We present a version of reliability described by Wanous et al.¹ for calculating single-item reliability. Single-item reliability can be estimated using the communality of a single item if the item can be embedded in a factor analysis of a larger set of items. As described by Wanous and Hardy (2001),² communality (variance item shares with a common factor) of an item as a conservative estimate of an items' reliability. The true reliability could be higher but cannot be lower.² However, for this method to be valid, we must first establish the candidate item performs reasonably in a factor analysis with other items. The care/communication factor was identified via principal axis factors analysis and is comprised of 6 BFS items. We included the BFS top-box overall rating in a principal axis factor analysis using polychoric correlations to address the violation of the normal item distribution

assumption.

-

References

1. Wanous JP, Reichers AE, Hudy MJ. Overall Job Satisfaction: How Good Are Single-item Measures? J Appl Psychol. 1997 Apr; 82(2):247-52.
2. Wanous JP, Hudy MJ. Single-Item Reliability: A Replication and Extension. Organizational Research Methods. Organ Res Methods. 2001. 4, 361-375. 10.1177/109442810144003.

2. Accountable Entity-Level

Accountable Entity-level Test-Retest Reliability Testing (Intraclass Correlation

Coefficient): To evaluate year-to-year reliability (stability) of the BFS-PM, we evaluated facility-level test-retest consistency with a two-way random effects ICC (2,1). The ICC(2,1) quantifies the strength of absolute agreement of scores between time1 (FY22) and time2 (FY23) at each facility.

Accountable Entity-Level Reliability Testing: We define a VA facility's true BFS-PM score as the observed percentage of respondents reporting BFS top-box overall scores of 9 or 10 on the 11-point Likert response scale (0=worst, 10=best). Reliability was assessed using the Adams beta-binomial approach¹ to calculate signal-to-noise ratios. Briefly, for each VA facility, we partition the two variance components comprising top-box BFS scores into between-facility variation (signal) and within-facility variation (noise). Conceptually, the ratio of signal/noise indicates the measure's ability distinguish between true facility-level differences in performance versus other sources of variation. The beta-binomial approach is appropriate for measures where each BFS respondent represents a binary opportunity for a facility to achieve a top-box overall rating of satisfaction. This signal-to-noise ratio ranges between 0 when all of the variability can be attributed to non-performance sources of variation (e.g., measurement error) and 1 when all the variability is due to true differences in quality of care across VA facilities. The within-facility variation component assumes sampling error in respondent scores follow a binomial distribution. The between-facility variation (true quality measurement) is assumed to follow a flexible beta distribution. Therefore, between-facility variation was estimated using a beta-binomial model via the BETABIN SAS macro (v2.2). The reliability estimate of the facility-level BFS-PM score for VA facility j is calculated as:

$$\text{reliability of facility } j = \frac{\text{Variance}_{\text{between-facility}}}{\text{Variance}_{\text{between-facility}} + \text{Variance}_{\text{within-facility-}j}}$$

References

1. Adams, J. L. (2009). The reliability of provider profiling: A tutorial. Santa Monica, CA: RAND Corporation.

5.2.3 Reliability Testing Results

Person-level Test-Retest Reliability Results Between BFS overall rating item scores from time1 and time2 (n=116)

Cohen's kappa¹ = 0.63 (0.45-0.81) --> Moderate agreement

ICC(2,1)² = 0.63 (0.51-0.73) --> Moderate agreement

Logistic regression (odds ratio)³ = 53.3 (95%CI: 10.7-264.8) --> Very strong association

Notes

-

¹Cohen suggested the Kappa result be interpreted as follows: values ≤ 0 as indicating no agreement and 0.01-0.20 as none to slight, 0.21-0.40 as fair, 0.41- 0.60 as moderate, 0.61-0.80 as substantial, and 0.81-1.00 as almost perfect agreement.

²Two-way random-effect Intraclass Correlation Coefficient (ICC) is a measurement of absolute agreement between test-retest of BFS item scores. ICCs from 0.4 - 0.6 are considered "moderate" agreement.

³The odds ratio represents the odds of reporting a BFS "top box" overall rating=1 at time2 based on BFS overall rating at time1, answering the question: how well does time1 BFS rating predict the time2 BFS rating?

-

Data Element Reliability Testing Results: Factor Analysis Approach to Evaluating Single-Item Reliability (n=5,812)

Based on Horn's¹ parallel analysis for determining the number of factors to retain, only 1-factor emerged with factor loadings ranging from 0.61– 0.83. The overall top-box rating item had a factor loading of 0.83, well-above the minimum recommended loading of 0.4.² The communality (reliability estimate) is 0.684, closely approximating the commonly used cutoff of 0.7 despite being known to underestimate the true reliability.³

-

References

1. Horn JL. A rationale and test for the number of factors in factor analysis. *Psychometrika*. 1965;30:179–185.
2. Nunnally JC. *Psychometric Theory*. New York, NY: McGraw-Hill (1978)
3. Nunnally J.C., Bernstein I.H. *Psychometric theory*. 3rd ed. New York, NY: McGraw-Hill, Inc.; 1994.

-

-

Facility-level Test-Retest Reliability Results for FY22 vs. FY23

The two-way random ICCs for facility-level test-retest for FY22 vs FY23 were as follows: no restrictions on the number of surveys: 0.59 (n=145), at least 5 surveys: 0.61 (n=133), at least 10 surveys: 0.64 (n=115), at least 20 surveys: 0.75 (n=96), and at least 30 surveys: 0.74 (n=67).

5.2.4 Interpretation of Reliability Results

The BFS top-box overall rating showed moderate to high reliability at the individual-level with a Cohen's kappa and ICC estimates of 0.63, and a very strong association between Time1 and Time2 ratings (OR = 53.3). The single-item reliability estimate via the Wanous et al.¹ factor analysis approach was 0.684, indicating good/excellent reliability.

At the facility level (accountable entity), year-to-year reliability as estimated by two-way random effects, the ICC ranged from 0.59 (no minimum number of respondents) to 0.75 for facilities with 20 or more respondents each year. This indicates moderate (0.59) to good reliability (≥ 0.7) at all sample sizes. Beta-binomial reliability scores ranged from 0.165 (2 facilities) to 0.99 (7 facilities) with a weighted mean reliability of 0.73. Values above 0.7 are considered acceptable.

¹Wanous JP, Reichers AE, Hudy MJ. Overall Job Satisfaction: How Good Are Single-item Measures? J Appl Psychol. 1997 Apr; 82(2):247-52.

Table 2. Accountable Entity Level Reliability Testing Results by Denominator, Target Population Size

Accountable Entity-Level Reliability Testing Results													
	Overall	Minimum	Decile_1	Decile_2	Decile_3	Decile_4	Decile_5	Decile_6	Decile_7	Decile_8	Decile_9	Decile_10	Maximum
Reliability	72.6	16.5	34.2	54.8	62.3	69.9	75.9	79.1	82.2	85.6	89.0	96.1	99.9
Mean													
Performance	77.4%	66.7%	71.3%	70.9%	68.9%	77.0%	73.6%	81.3%	78.1%	78.6%	81.6%	92.6%	100.0%
Score													
N of Entities	145	2	16	13	15	14	15	14	15	14	15	14	7
N of Persons													
/ Encounters	5,812	6	111	206	339	371	604	522	791	937	1,181	750	35
/ Episodes													

5.3.1 Level(s) of Validity Testing Conducted

Person or encounter level (i.e., data element) (e.g., sensitivity and specificity), Accountable entity level (i.e., measure score) (e.g., criterion validity)

5.3.3 Method(s) of Validity Testing

Construct Validity (Individual and facility-level)

-

Construct validity was evaluated at the individual-level (using logistic regression) and the facility-level (using Pearson correlations) and examined associations between the BFS top-box overall rating and theoretically related care processes (i.e., quality of care indicators) obtained from the scientific literature (see section 2.1 Logic Model). We have hypothesized that at the individual-level, receipt of each of these quality indicators or “Best Practices” should result in a statistically significant higher BFS overall rating, and at the facility-level, that higher proportions of Veterans/families receiving these indicators would be associated with higher BFS-PM scores.

We developed several chart-derived process variables based on the empirical literature and “Best Practices” as outlined in the National Consensus Project for Quality Palliative Care Clinical Guideline. These variables included: 1) receipt of a comprehensive palliative consult in the patient’s last 90 days of life; 2) death in a dedicated inpatient hospice unit; 3) patient or family contact with a chaplain in the last month of life; and 4) evidence of emotional support given to a family member up to two weeks post-Veteran death [bereavement contact]. All process variables were dichotomized (receipt/no receipt) for individual-level analysis. For the facility-level analysis, the proportion of Veterans/families who received each indicator in a facility was calculated.

We also examined associations between the BFS “top box” overall rating and a set of other EOL

care processes derived from individual BFS items. For the individual-level analysis, individual BFS items were dichotomized to reflect the most optimal response (“always”) vs. all others. For the facility-level analysis, optimal responses for the individual BFS items were aggregated to create facility-level proportions.

Validation of process measure data collection: Currently, all variables are extracted directly from the VA Corporate Data Warehouse (an integrated system of national databases including clinical, administrative and financial data) using standardized algorithms.

Nonresponse Bias Adjustments

Prior to reporting, facility-level scores are weighted to account for survey non-response. Multivariable logistic regression was used to generate the adjusted predicted probability of a completed BFS. Weights are generated by taking the inverse of the adjusted predicted probability. The nonresponse model includes the following variables: post-death bereavement contact of a family member, receipt of chaplain visit/contact, receipt of a palliative care consultation, death in community living center (i.e. a VA-operated nursing home) or inpatient hospice unit, next-of-kin (BFS respondent) relationship, Veteran race and ethnicity, Veteran age, Veteran sex, length of terminal admission (in days), a set of 38 individual comorbid health conditions (e.g. obesity, congestive heart failure) and total count of comorbid conditions as defined by Elixhauser and colleagues.¹ Weights are re-calculated quarterly.

-

References

¹ Agency for Healthcare Research & Quality. (2023) Elixhauser Comorbidity Software Refined for ICD-10-CM Diagnoses, v. 2024.1. Available from <https://hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/CMR-User-Guide-...>

5.3.4 Validity Testing Results

Construct Validity: Individual-level

At the individual-level, we calculated unadjusted odds ratios (ORs) and 95% confidence intervals (CI) to demonstrate the association between receipt of quality of end-of-life care process measures (derived from the Veteran’s medical record and individual items on the BFS), and a BFS overall rating “top box” score.

Chart-derived process measures (n=11,568)

Receipt of a palliative care consult: OR=1.94, 95% CI: 1.74-2.14, $p<0.001$

Death in a dedicated inpatient hospice unit: OR=2.19, 95%CI:2.00-2.40, $p<0.001$

Chaplain contact before death: OR=1.38, 95%CI:1.21-1.58, $p<0.001$

Bereavement contact after death: OR=1.29, 95%CI: 1.18-1.42, $p<0.001$

Individual BFS item process measures (n=11,568)

-

Doctor and staff listened to concerns: OR=17.93, 95%CI:16.11-19.97, $p<0.001$

Doctors and staff were kind, caring and respectful: OR=19.66, 95%CI:17.34-22.28, $p<0.001$

Personal care needs were met: OR=11.82, 95%CI: 10.68-13.09, $p<0.001$

Pain was well-managed: OR=10.44, 95%CI:9.44-11.55, $p<0.001$

Dyspnea was well managed: OR=11.68, 95%CI:10.41-13.09, $p<0.001$

Emotional support before death needs were met: OR=13.74, 95%CI:12.38-15.25, $p<0.001$

Construct Validity: Facility-level (n=145)

Pearson correlations were used to evaluate the unadjusted associations between the BFS-PM and established chart-derived EOL care process measures at the facility-level. The BFS-PM exhibited a positive correlation with the proportion of patients who received a palliative care consultation with a moderate effect size ($r=0.2831$, $p<0.05$) and a positive correlation with the proportion of patients who died on a dedicated inpatient hospice unit with a strong effect size ($r=0.4934$, $p<0.05$). Positive, but statistically insignificant correlations were observed between the BFS-PM and receipt of a chaplain contact ($r=0.0608$) and receipt of a post-death bereavement contact ($r=0.1080$).

Nonresponse bias model (n=14,510):

The R^2 for the nonresponse bias adjustment model was 0.049 with an area under the curve (AUC) of 0.6486. The change in facility-level scores from pre- to post-weighting for nonresponse ranged from -0.110 to 0.107. Please see **section 4.3.4a (Additional Validity Testing Results attachment)**. The attached zip file includes an Excel (.xls) file of facility-level BFS-PM scores before and after weighting for nonresponse (first tab). The zip file also includes a Word document with the individual- and facility-level distributions of the nonresponse model variables (Table 1, p. 1-2), and final nonresponse model specifications and fit statistics (Table 2, p. 3-4 and Figure 1, p. 4).

5.3.4a Attach Additional Validity Testing Results

[4.3.4a. Additional Validity Testing Results.zip](#)

5.3.5 Interpretation of Validity Results

The BFS-PM (and the BFS global rating item) has shown, through peer-reviewed analyses, its ability to distinguish among groups that should, in theory, have different scores. The BFS-PM meets the standards of construct validity. All process measures at the individual- and facility-levels were positively associated with the BFS-PM (and the BFS global rating item) and demonstrated at least marginal statistical significance (most were highly significant).

5.3.2 Type of Accountable Entity Level Validity Testing Conducted (derived)

Empirical validity testing at the accountable entity-level (e.g., criterion validity, construct validity, known groups analysis), Systematic assessment of face validity of the measure's performance score as an indicator of quality or resource use

5.4.1 Methods Used to Address Risk Factors

Statistical risk adjustment model with risk factors

5.4.2 Conceptual Model Rationale

Prior to reporting, BFS-PM scores are adjusted for nonresponse (see section 4.3.3.) and case-mix.

Our risk adjustment approach for facility case-mix is informed by the scientific literature, including models used by other established measures of family-reported quality of end-of-life (EOL) care, and our own published and unpublished internal evaluations. The attached conceptual model (see section 4.4.2a) outlines all risk factors that we considered for inclusion in our risk adjustment model for facility case-mix.

In the top portion of the figure, the blue shaded box includes factors related to the quality of care provided by the accountable entity (i.e., VA facilities) including provider practices, potential biases and/or discrimination, and facility characteristics (e.g., size, teaching affiliation). While these factors are associated with EOL care processes (e.g. receipt of a palliative care consultation) that influence the BFS overall rating (and thus, the BFS-PM), they should not be used as risk-adjustment variables. As these factors are presumably under the facility's control, facilities are responsible for addressing the needs of the patient populations they serve through changes in care process delivery.

The bottom portion of the figure includes factors potentially influencing the BFS overall rating that are present at the start of care (i.e., patient and respondent characteristics), but independent of the quality of EOL care received. As a general rule, patient and respondent characteristics that are not under the facility's control, but are associated with bereaved family member ratings of quality of EOL care (i.e., the BFS global rating), should be considered for inclusion in a case-mix adjustment model.¹ These domains (noted with brackets) include health status and clinical conditions, patient and respondent demographics, comorbid conditions, social vulnerabilities, lived environment, and functional factors.

The case-mix adjustment model for the BFS-PM includes the following variables (indicated in the figure using *italics*): Veteran's age at the time of death; a set of 38 individual comorbid health conditions (e.g. obesity, congestive heart failure) and total count of comorbid conditions as defined by Elixhauser and colleagues² diagnosed in the two years prior to, and including, the terminal admission; Veteran's primary diagnosis on terminal admission; BFS respondent's relationship to the Veteran (e.g., spouse), respondent's highest level of education (e.g. 4-year college graduate), and mode of administration (e.g., mail).

The current model was informed by our original case-mix adjustment approach³ developed in 2016 and the case-mix adjustment model employed by the Centers for Medicare & Medicaid Services (CMS) Consumer Assessment of Healthcare Providers and Systems (CAHPS) Hospice survey.⁴ Beyond the CAHPS Hospice survey, there is limited information in the literature on the adjustment of differences in case-mix for family-reported EOL care ratings. Per the most recent case-mix adjustment documentation available from CMS, the CAHPS Hospice case-mix adjustment model includes the following variables: mode of survey administration, response percentile, decedent age, payer, primary diagnosis, length of hospice stay, caregiver age, caregiver

education, caregiver relationship to decedent and language spoken at home.⁴ All relevant CAHPS-Hospice survey risk-adjustment variables that were available in VA databases were included in creating the BFS case-mix adjustment model, including decedent age, primary diagnosis, and respondent's highest level of education and relationship to the decedent.

Additional CAHPS Hospice variables (i.e, primary language spoken in the home, caregiver age) were considered but were not available to us.

Our approach extends the CAHPS Hospice model and includes adjustment for the total number of comorbidities experienced by the Veteran during the last two years of life. Comorbid health conditions were found to have large and statistically significant effects on BFS-PM scores (and facility rankings) in our prior case-mix adjustment analyses.³ Further, we include this additional clinical risk factor to align with the Agency for Healthcare Research & Quality (AHRQ)'s recommendation to adjust for patient severity of illness when making facility-level quality comparisons.⁵ Finally, we include adjustment for mode of BFS administration (e.g. mail, phone).⁶

In alignment with CAHPS Hospice case-mix adjustment and AHRQ guidance, we do not include social and functional risk factors (e.g., race/ethnicity, income, lived environment, functional status) in our case-mix adjustment model. When adjusting healthcare quality scores for the purposes of comparison across providers, AHRQ states that "most common risk adjustments—for age, prior medical history, or comorbidities—are considered a good idea. It is not considered appropriate, in most circumstances, to adjust for other sociodemographic characteristics such as race, ethnicity, income, and/or insurance status. Such adjustments may essentially conceal unacceptable disparities in care."⁵

References

1. Elliott MN, Zaslavsky AM, Goldstein E, Lehrman W, Hambarsoomians K, Beckett MK, Giordano L. Effects of survey mode, patient mix, and nonresponse on CAHPS hospital survey scores. *Health Serv Res*. 2009 Apr;44(2 Pt 1):501-18. doi: 10.1111/j.1475-6773.2008.00914.x. PMID: 19317857; PMCID: PMC2677051
2. Agency for Healthcare Research & Quality. (2023) Elixhauser Comorbidity Software Refined for ICD-10-CM Diagnoses, v. 2024.1. Available from <https://hcup-us.ahrq.gov/toolssoftware/comorbidityicd10/CMR-User-Guide-...>
3. Kutney-Lee A, Carpenter J, Smith D, Thorpe J, Tudose A, Ersek M. Case-Mix Adjustment of the

Bereaved Family Survey. Am J Hosp Palliat Care. 2018 Jul;35(7):1015-1022.
doi: 10.1177/1049909117752669.

4. CAHPS Hospice Survey. Scoring and Analysis. Available from:
<https://hospicecahpsurvey.org/en/public-reporting/scoring-and-analysis/>.

5. Agency for Healthcare Research and Quality. Making Adjustments to Health Care Quality Scores. Available from <https://www.ahrq.gov/talkingquality/translate/scores/adjustment-scoring...>

6. Thorpe JM, Smith D, Kuzla N, Scott L, Ersek M. Does Mode of Survey Administration Matter? Using Measurement Invariance to Validate the Mail and Telephone Versions of the Bereaved Family Survey. J Pain Symptom Manage. 2016 Mar;51(3):546-56. doi: 10.1016/j.jpainsymman.2015.11.006.

5.4.2a Attach Conceptual Model

[4.4.2a. BFS_Case_Mix_Adjustment_figure_final.pdf](#)

5.4.3 Variable Distribution Across Measured Entities

Due to the number of variables, the distribution of risk factor characteristics at the patient- and facility-levels is presented in **section 7.1., Supplemental Appendix B, Table 3 (p.3-4)**.

5.4.4 Risk/Case-Mix Adjustment Modeling and/or Stratification Results

VA uses risk adjustment weights to account for individual- and facility-level differences on factors other than quality of care, such as differences in patient medical complexity, setting of death (e.g., ICU vs. VA nursing home) and next-of-kin relationship to the Veteran. Multivariable logistic regression was used to generate the adjusted predicted probability of reporting a top-box BFS overall rating. Risk adjustment weights are generated by taking the inverse of the adjusted predicted probability. Weights are re-calculated quarterly. See final risk adjustment model specifications in attached Excel (.xls) file in section 4.4.4a.

The R2 for the case mix adjustment model was 0.036 with an area under the curve (AUC) of 0.6012. The change in facility-level scores from pre- to post-weighting for case-mix ranged from -0.014 to 0.047. See **section 4.4.5a (Calibration and Discrimination Testing Results attachment)**. The attached zip file includes an Excel (.xls) file of facility-level BFS-PM scores before and after weighting for case-mix (second tab). Results of Calibration and Discrimination testing for the overall and stratified risk adjustment models are presented in section 4.4.5 and 4.4.5a.

5.4.4a Attach Risk/Case-mix Adjustment Modeling and/or Stratification Specifications

[4.4.4a. Final Risk Adjustment Modeling Specifications_1NOV.xlsx](#)

5.4.5 Calibration and Discrimination

BFS data from FY22 (N = 5,812) were randomly partitioned into a 70% training set (n = 4,068) and a 30% hold-out test set (n = 1,744) for discrimination and calibration testing. Guided by the conceptual model, multivariable logistic regression using the training dataset was used to generate case-mix adjusted predicted probability of top-box BFS overall rating. Discrimination and calibration estimates are based on the 30% hold-out testing data.

In addition to providing VA facilities with BFS-PM scores, we also provide results stratified by venue of death (acute hospital/ICU vs. nursing home/inpatient hospice). Therefore, we report overall (pooled) discrimination/calibration results followed by stratified results.

Overall BFS Discrimination and Calibration Results (pooled across venue of death)

Discrimination statistics

1. **C-statistic:** Area under the receiver-operator characteristic (AUC) curve (the c-statistic) describes the probability that a randomly selected BFS respondent who reported a top-box BFS overall rating had a higher expected value than a randomly selected respondent who reported a lower BFS rating. Cross-validated mean AUC of 0.60 (bootstrapped 95% CI: 0.56 - 0.63) indicates fair model discrimination. This c-statistic is likely low because satisfaction with care is less determined by patient comorbidities than other outcomes. See **section 4.4.5a.(Calibration and Discrimination Testing Results attachment, Word document, Figure 1, p.1)** for a plot of the cvAUC and k-fold ROC curves.

Calibration

1. **Hosmer-Lemeshow goodness of fit (HL-GOF).** HL-GOF is a chi-square assessment of goodness of model fit by comparing the observed event rates to expected event rates across deciles of predicted proportions reporting the top-box BFS overall rating. The null hypothesis is there is no difference in observed/expected rates across deciles. The associated p-value based on the hold-out test data was 0.99 (HL chi2 (df, 8) = 2.06).

2. **Calibration belts.** Calibration was assessed using calibration belts as recommended by Nattino et al (2017).¹ Calibration belts plot the observed against the predicted probability of a top-box BFS overall rating across the entire prediction range. Calibration belts also generate calibration-in-the large (CITL) and calibration slope. CITL tells us whether the predicted ratings are overestimated or underestimated, on average. A calibration slope of 1.0 reflects perfect prediction of a top-box BFS rating, whereas a slope < 1 indicates average underestimation. Finally, calibration belts provide an overall statistical goodness of fit test of the null hypothesis that prediction accuracy is comparable across the entire prediction range. See results presented in **section 4.4.5a. (Calibration and Discrimination Testing Results attachment, Word document, Figure 2, p.2)**. Observed vs expected probabilities track closely to the slope of 1.0 (null hypothesis: no difference) with all bin prediction 95% CIs crossing the null hypothesis slope. The overall p-value based on the hold-out test data was 0.68 (NS at alpha=0.05), indicating model predictions were consistent across the prediction range.

Stratified BFS Discrimination and Calibration Results by Venue of Death

Discrimination

1. We compared ROC-AUC values across venue of death (acute care/ICU vs. nursing home/inpatient hospice). See **section 4.4.5a. (Calibration and Discrimination Testing Results attachment, Word document, Figure 3, p.3)**. The AUC did not differ significant across settings (chi2[1] = 0.38; p=0.539).

Calibration

1. We compared calibration belts across venues of death. Model predictions did not differ significantly across the prediction range for the two groups (acute/ICU, p=0.59; nursing home/inpatient hospice, p=0.59). See **section 4.4.5a. (Calibration and Discrimination Testing Results attachment, Word document, Figure 3, p.3)**.

-

References

Nattino, G., Lemeshow, S., Phillips, G., Finazzi, S., & Bertolini, G. (2017). Assessing the Calibration of Dichotomous Outcome Models with the Calibration Belt. The Stata Journal, 17(4),

1003-1014.

5.4.5a Attach Calibration and Discrimination Testing Results

[4.4.5a. Calibration and Discrimination Testing Results.zip](#)

5.4.6 Interpretation of Risk/Case-mix Factor Findings

The final risk-prediction logistic regression model identified statistically significant differences in the probability of reporting a top-box BFS overall rating across 16 different predictors. The models showed modest discrimination (C-statistic = 0.60) and was well-calibrated across the range of predicted outcomes. Taken together with the information on discrimination and calibration provided in the aforementioned sections, results demonstrate the final risk-adjustment adequately adjusts for characteristics identified in the guiding conceptual model.

Ultimately, we combine the nonresponse and case-mix weights (nonresponse weight * case-mix weight) and apply to facility-level BFS-PM scores for reporting. Changes in facility-level BFS scores after applying the combined weight ranged from -0.102 to 0.13 (see **section 4.4.5a (Calibration and Discrimination Testing Results attachment)**). The attached zip file includes an Excel (.xls) file of facility-level BFS-PM scores before and after applying the combined weight (third tab).

5.4.7 Final Approach to Address Risk Factors

Statistical risk adjustment model with risk factors

6.1.1 Current Status

In use

6.1.3 Current Use(s)

Public Reporting, Quality Improvement (Internal to the specific organization)

6.1.3 Program Details

Name of the program and sponsor

National Hospice and Palliative Care Program - Geriatrics and Extended Care, Veterans Health Administration, Department of Veteran Affairs

URL of the program

https://www.cherp.research.va.gov/PROMISE/The_Bereaved_Family_Survey.asp ch.va...

Purpose of the program

To improve the quality of end of life care for seriously ill Veterans and their families and to reduce variability in the quality of end-of-life care received in VA Medical Centers across the nation.

Geographic area and percentage of accountable entities and patients included

United States of America, including Puerto Rico, Alaska and Hawaii

Applicable level of analysis and care setting

Patient-level, facility-level. All inpatient settings including acute venues, community living centers

and inpatient hospice units.

6.2.1 Actions of Measured Entities to Improve Performance

VA's Veteran Experience Center along with the Hospice and Palliative Care Implementation Center have regular (and as-needed) contact with users to assist with interpreting BFS results which are updated quarterly and planning performance improvement. A Compendium of Best Practices SharePoint site is available to assist facilities with selecting targeted interventions based on BFS data analysis and BFS-PM trend data is provided to track the impact of quality improvement efforts over time. Facilities are asked to create yearly S.M.A.R.T. (specific, measurable, achievable, relevant and time-limited) goals related to their palliative and hospice care program. Goals are documented and monitored for progress and achievement.

6.2.2 Feedback on Measure Performance

Quarterly reports, including the BFS-PM at the national and facility-level, and process-based quality indicators obtained from the BFS and patient medical records, are provided to VA leadership, policymakers, and VA hospice and palliative care teams located at each VA facility across the country. Users have been very accepting of the survey's findings and are open to suggestions for increasing BFS-PM scores. Feedback on the measure and its performance is obtained through regular conference calls with users. Finally, in a recent internal survey of representatives of VA Palliative and Hospice Care teams nationally, 83.5% of respondents reported that the BFS was "somewhat" or "very" important to facility leadership.

6.2.3 Consideration of Measure Feedback

In response to prior feedback obtained from users, we have integrated non-response bias and case-mix weights to BFS-PM scores prior to reporting to facilitate fairer comparisons.

6.2.4 Progress on Improvement

The national BFS-PM has increased over time: 71% (2021), 73% (2022), 78% (2023), and 80% in 2024 (through September 30). However, significant variation remains across facilities. For the time period of October 1, 2021, to September 30, 2023, facility-level BFS-PM scores ranged from 55% to 100%.

6.2.5 Unexpected Findings

In the process of collecting the BFS, several unintended impacts have been identified, such as suicidality among bereaved family members and unmet needs related to bereavement services and VA benefits. Staff at the Veteran Experience Center are now trained and able to assist family members with each of these issues and make appropriate referrals.

7.1 Supplemental Attachment

[7.1 Supplemental Attachment.zip](#)

Developer POC email

ann.lee@va.gov

Measure Developer POC

Ann Kutney-Lee
Veteran Experience Center, Department of Veteran Affairs
Philadelphia, PA
United States

The measure developer is different from the measure steward

No

Steward Address

Dawn Gilbert
Philadelphia, PA
United States

Steward Organization

Department of Veterans Affairs / Hospice and Palliative Care

Steward Organization URL

https://www.va.gov/GERIATRICS/index.asp?utm_source=geriatrics_left_menu

Steward POC email

dawn.gilbert2@va.gov