Peer Review File

Article information: https://dx.doi.org/10.21037/apm-23-508

Reviewer A:

Comment: Although the article is described as a systematic review of behavioral health interventions, it includes pharmacotherapy and is a study on behavioral health comorbidities.

Response: We appreciate the reviewer's concern about the lack of clarity in our terminology. We have changed the language in the title and abstract to reflect the concern raised by reviewer A. We define behavioral health comorbidities in the second paragraph of the introduction, however, we have also now added a clarification to section 1.3 defining behavioral health interventions to prevent confusion with behavioral interventions.

Changes in Text:

Line 1-2: Title changed to "Interventions for Behavioral Health Comorbidities in the Hospice Setting: A Scoping Review."

Line 45-46: Language changed from "behavioral health intervention studies" to "studies of interventions for behavioral health comorbidities in the hospice setting."

Highlight box: Changed "behavioral health interventions" to "interventions targeting behavioral health conditions" and "interventions specifically for behavioral health."

Highlight box: Changed the implications sentence to "A greater emphasis must be placed on building infrastructure to successfully conduct research on behavioral health comorbidities in hospice settings, with a focus on generating implementation-ready, scalable interventions appropriate to patients at the end-of-life."

Line 119: Changed "behavioral health interventions" to "interventions targeting behavioral health conditions."

Line 130: "Behavioral health interventions" changed to "mental health treatments."

Line 141: We added the following to clarify our use of the term "behavioral health intervention" for the remainder of the manuscript: "We define behavioral health interventions as interventions specifically targeting mental health and substance use disorders or psychiatric symptoms (e.g., depression, anxiety) as their primary therapeutic target."

Comment: However, if it is a study on intervention on behavioral health comorbidities, there would be numerous studies on depression, anxiety, etc., and I do not think it has been sufficiently searched. Palliative care units may also play a similar role to hospice, so it may not be sufficient to limit the search to the term "hospice." What is the intention of focusing on the hospice setting? Isn't the characteristics of the patient important, not the setting.

Response: We chose to focus on hospice settings because in our context (the United States), hospice services are highly distinct from other palliative care delivery structures. In the United States, patients who elect to enroll in hospice generally cease receiving disease-oriented treatment from any other sources and often transition to having their palliative care needs met through hospice service delivery rather than through other forms of palliative care (e.g., hospital-based ambulatory or inpatient palliative care services).. There are unique structural features of hospice care delivery in the United States that may be highly impactful to care delivery relative to other palliative care services. For example, palliative care clinicians practicing in an inpatient acute care/hospital setting often have access to psychiatric consultation. However, hospice staffing models do not typically include a full range of medical services like hospital-based care. Unlike other forms of palliative care in the United States, most hospice care is delivered in the home setting.

We are aware of many additional studies conducted in end-of-life or palliative care settings focusing on depression or anxiety, but we elected not to include these because of the significant health services and care delivery differences involved in hospice care in our context. We hope to have provided reasoning for our initial decision to focus on hospice services. However, one of the major limitations that we came across in conducting our study is the heterogeneity in language around palliative care and hospice services in different contexts.

Changes in Text:

In section 2.3, subsection Data synthesis, we added the following:

While analyzing data, we recognized that heterogeneity in the use of the term hospice between health systems presented an unexpected challenge. While some health systems (e.g., those of the United States and the United Kingdom) distinguish the hospice model of care from other palliative care services, this distinction is not universally recognized. Furthermore, the definition of hospice varies across health systems. To avoid excluding potentially relevant studies but also to highlight research in hospice as a unique care delivery model, we present data for all identified articles and also present subgroup analyses of studies conducted in the United Kingdom and Ireland (32–34), Australia (35), New Zealand (36), the United States (37), Canada (38), Ireland, and Poland (39). These countries were included because they utilize hospice models consisting of distinct structures of care, at least some of which are delivered in the community/home setting and targeted towards individuals at the end of life. However, even among these systems, there is significant variability in how care is delivered at home versus facilities, regulations regarding concurrent disease-modifying treatments, distinct hospice providers, and payment models. We subsequently refer to these as *the hospice model subgroup*.

We excluded other countries from the subgroup analysis if the model of hospice appeared to consist predominantly of inpatient palliative care wards/units (e.g., China (40), Japan (41,42)), if there seemed to be no distinction between hospice and palliative care services (e.g., India (43), Bosnia (44)), or if we were unable to clarify the model of care within a given health system.

In section 3.2 we added the following:

Among hospice model subgroup studies, fewer (25%) of studies were predominantly inpatient. In contrast, most studies conducted in health systems that deliver only inpatient hospice or do not distinguish between palliative care and hospice (67%) were conducted in inpatient settings.

&

Sample sizes ranged from 11 to 629 participants (mean 82.9, standard deviation 122.9). This was similar among the hospice model subgroup studies (mean 97.7; standard deviation 136.4).

In section 3.3 we added the following:

The most common types of interventions involved the use of complementary alternative medicine, such as massage, acupuncture, music therapy, and yoga (35% of total studies; 36% of hospice model subgroup studies) rather than conventional behavioral health interventions.

&

Of note, all pharmacotherapy studies were conducted in the hospice model subgroup.

&

We added hospice model subgroup data to key study characteristics regarding measurement and outcome: Table 3 presents the analyzed studies' behavioral outcomes (BH) outcomes. Studies universally measured symptoms rather than diagnoses (e.g., depressive symptoms rather than major depressive disorder). The most commonly employed measures used to assess treatment effect included the Hospital Anxiety and Depression Scale (HADS) (in 24% of all studies and 32% of hospice model subgroup studies) (80), the Edmonton Symptom Assessment System-Revised (ESAS) (16% total; 14% of hospice model subgroup studies) (81), the Clinical Global Impression Scale (CGI) (8%) (82), and the Patient Health Questionnaire (PHQ) (8%) (83). Depressive symptoms (81% of all studies; 79% of hospice model subgroup studies) and anxiety symptoms (59% of all studies; 57% of hospice model subgroup studies) were the most prevalent behavioral health outcomes. Other constructs measured included distress (n=5), mood (n=3), and well-being (n=4). As a note, we distinguished mood from depression scales because the three studies that used mood as an outcome utilized subjective mood scales that asked about a respondent's overall mood rather than about depression explicitly (e.g., the Memorial Pain Assessment Card asks respondents to rate themselves from “worst mood” to “best mood”) (84). A minority of studies (11%, both hospice model subgroup) employed unvalidated scales or were not specific in their measurement approach. No studies focused on substance use disorders, post-traumatic stress disorder, or serious mental illnesses like psychotic disorders.

In section 3.4 we added the following:

Among the 27 studies in the hospice model subgroup, 9 (33%) had statistically significant results.

In the discussion, we added the following paragraph on the challenges of reviewing hospice-based studies from across health systems.

Our study contributes to the field of hospice research, focusing specifically on hospice care settings rather than all palliative care more broadly. We identify key research gaps with direct implications for current clinical practice. However, our study has several limitations. The *a priori* criteria we established in our protocols introduced significant heterogeneity into our study, and thus, our findings must be couched in a broader process-level challenge. Our goal in conducting this study—to explore the landscape of behavioral health research in the hospice setting—was informed by our context as researchers and clinicians in the United States. Hospice care in the United States is a distinct model of care with specific payment and care delivery structures that distinguish it from other forms of palliative and end-of-life care. As such, we elected to limit our search and inclusion criteria to hospice and not include other palliative and end-of-life care models. We specifically chose to privilege the care setting (hospice) over the population (patients at the end of life) because of the distinct challenges of conducting research and implementing interventions in the (U.S.) hospice setting. In conducting our study, we found that much of the terminology that we used did not translate into other health systems. Many health systems interchangeably use language such as palliative care, hospice, and end-of-life care. Furthermore, among systems that distinguish hospice as a specific model of care, the definition of the model may vary significantly and range from a blanket term for palliative care to a model of predominantly in-patient end-of-life care to largely community-based end-of-life care. Many of the nuances, such as the possibility of receiving hospice care concurrently with disease-oriented treatments (which is generally not allowed in United States hospice care for adults), differ between settings. As a result, we may have excluded studies that were conducted in applicable settings, perhaps using different terminology, and we included studies that do not reflect the hospice context in which we were interested. We attempted to mitigate this challenge by specifying those studies conducted in the United States and those undertaken in systems with similar hospice care structures (the hospice model subgroup); this presents a major limitation and an important process-level finding of our study. In attempting to address this weakness in our study, we were surprised to find the overall shortage of descriptions of hospice as a care model across settings. In this sense, a key recommendation emerging from our study is the need for literature describing models of hospice care across health settings to ensure that researchers conducting studies such as ours oriented around a health delivery model can appropriately include studies conducted globally.

Reviewer B:

Comment: "Relevant vocabulary for concepts including hospice care and mental/behavioral health" - Would be beneficial to reference Appendix 1 at this point.
Response: Thank you for this suggestion. This has been done.
Changes in Text: *Section 2.1:* "(see Appendix 1)" *added as suggested.*

Comment: Population - unclear regarding statements of 33% - does this refer to the study population needing to be at least 33% hospice patients, or 33% adult patients? Was this done to have comparison? If only 33% of the study is done with hospice patients is it appropriate to suggest that this is a study of hospice patients? Would be better to standardize terminology - "at least 33%", and then later ">33% (with underline under >)"
Response: We agree with the reviewer about this lack of clarity and have made the suggested revision.

Changes in Text: *The population subsection of manuscript section 2.2 now reads:* For studies conducted outside the U.S., we also included studies in which at least 33% of patients received care at a hospice-equivalent end-of-life care paradigm (e.g., a care setting providing care focused on quality of life to patients with a prognosis of six months or less).

Comment: You mention that caregiver focused studies were excluded in 2.2, but later mentioned that they were included in non-intervention studies
Response: Thank you for pointing out that we failed to adequately clarify the status of caregiver-focused studies. This has been revised in section 2.3 of the text.
Changes in Text: *Section 2.3 of the text now reads:* Studies screened for full-text review were initially designated as intervention or non-intervention studies. Given that this study included only patient-focused interventions, studies focused on caregivers were excluded from the present review. Two independent study team members evaluated full-text intervention studies with discrepancies resolved by senior author DS.

Comment: Line 285 - when mentioning that depression and anxiety are the most prevalent outcomes it may be beneficial to state depressive symptoms/anxiety symptoms instead to prevent confusion that these were diagnosed disorders, similarly to line 280 and 281.
Response: We agree with the reviewer's comment and have changed the text accordingly.
Changes in Text: *In section 3.4, the section commented on by the reviewer now reads:* Depressive symptoms (81% of all studies; 79% of hospice model subgroup studies) and anxiety symptoms (59% of all studies; 57% of hospice model subgroup studies) were the most prevalent behavioral health outcomes.

Comment: 286 - how does mood outcome differentiate from depression?
Response: We appreciate the reviewer pointing out the confusion in these constructs. We have clarified our reasoning for distinguishing mood and depression as outcomes in section 3.3.
Changes in Text: *Section 3.3 now reads:* Other constructs measured included distress (n=5), mood (n=3), and well-being (n=4). As a note, we distinguished mood from depression scales because the three studies that used mood as an outcome utilized subjective mood scales that asked about a respondent's overall mood rather than about depression explicitly (e.g., the Memorial Pain Assessment Card asks respondents to rate themselves from "worst mood" to "best mood") (84).

Comment: 3.4 first chapter - again when saying anxiety and depression are you referring to symptoms or disorders? ie. does depression just mean low mood or does it refer to a full depressive episode? May be beneficial to have a line earlier in paper to advise that depression will be used interchangeably with depressive symptoms and clarify what this entails.

Response: We agree with the reviewer regarding this clarification. We have added in "symptoms" as a clarification where discussion changes in anxiety and depression symptoms in section 3.4.

Changes in Text: *Section 3.4 now reads:*

Notable study outcomes are described in Table 3. Slightly fewer than half (17 studies; 46%) of studies demonstrated statistically significant findings related to a behavioral health outcome (i.e., reduction in depression/anxiety symptoms over time between groups) (45,54–56,58,61,63,64,68,70–74,77,78,85). Among the 27 studies in the hospice model subgroup, 9 (33%) had statistically significant results.

Among those studies demonstrating significant results, four significantly improved only anxiety symptoms (45,54,70,77). Two of these studies were CAM interventions, and two were psychotherapy interventions. Two studies significantly improved only depression symptoms (58,85): one CAM study and one psychotherapy intervention. Six studies significantly improved both anxiety and depression symptoms (56,64,71–73,78). Interventions that improved both anxiety and depression symptoms were CAM (3), spiritual care/other (1), combined psychotherapy and CAM (1), and psychotherapy 1. The remaining five studies improved mood, distress, or overall clinical impression (55,61,63,68,74). These studies consisted of CAM (3), pharmacologic (1), and psychotherapy (1) interventions. Overall, Among the 12 studies that significantly improved depression and/or anxiety symptoms, four employed psychotherapy interventions (45,54,78,85), six employed complementary and alternative medicine interventions (56,58,64,70,71,77), one employed combined psychotherapy and complementary and alternative intervention (72), and one employed spiritual care (73).

Among studies that provided a magnitude of change, improvements in depression and anxiety symptoms were relatively modest. For instance, among the two studies utilizing the HADS and demonstrating significant improvements in depression and anxiety symptoms, the magnitude of change on the HADS subscales ranged from 2.7 to 4.7 points (56,78). Of note, these values exceed the proposed minimally important difference of 1.5 points on each component of the HADS (derived from patients with cardiovascular and pulmonary disease) (86,87). Similarly, studies utilizing the ESAS and demonstrating significant improvements in depression and anxiety symptom scores ranged from 1.0 to 2.3 points change on a 10-point scale (64,77). These changes overlap and may exceed the minimally significant difference in individual ESAS symptom measures (88). Most studies measured improvements immediately after the intervention or study, and few demonstrated sustained improvement in symptoms over time following the intervention.

Comment: 3.4 HADS, ESAS - used for first time without introducing the abbreviations earlier in paper. May be in chart however best to also include this in the written manuscript in line 283 for both.

Response: This has been addressed in section 3.3.

Changes in Text: *In section 3.3:* The most commonly employed measures used to assess treatment effect included the Hospital Anxiety and Depression Scale (HADS) (in

24% of all studies and 32% of hospice model subgroup studies) (80), the Edmonton Symptom Assessment System-Revised (ESAS) (16% total; 14% of hospice model subgroup studies) (81), the Clinical Global Impression Scale (CGI) (8%) (82), and the Patient Health Questionnaire (PHQ) (8%) (83).

Comment: Line 347 - suggest changing country to region as the United Kingdom is a union of 4 countries, rather than 1 single country.
Response: We appreciate the reviewer identifying this oversight and have amended both our tables and the text.
Changes in Text: *In the discussion, we now state:* For instance, the most represented region in our study was the United Kingdom, where a large proportion of hospice services are provided in inpatient or residential settings (93).

Comment: I have published research on this topic, which you have cited in your paper. I note that many of the articles I cited did not come up in your review. I wonder if this is due to the search terms used. It would be beneficial to branch out your review to use terms such as end of life care or palliative, rather than just hospice as this may not be a commonly used term globally.
Response: We appreciate the reviewer's important point here—one which highlights a key limitation of our study. The reviewer's comment aligns with reviewer A's second comment, which we attempted to address through a range of alterations and additions to the text, including the creation of the hospice model subgroup and the additional discussion of this limitation in the discussion. We will refrain from highlighting these changes in the text again, however, we wish to reinforce that we very much recognize the reviewer's critique here. In a sense, we hope to highlight that in addition to the findings of the study itself, the methodological barrier that we encountered in translating the terminology of our care context to the global delivery of palliative care is a key finding of our paper. To that end, in addition to the text changes we highlighted in our response to Reviewer A, we have also addended our conclusion to reflect this limitation. We would also be happy to continue revising to ensure we are inclusive of appropriate studies.
Changes in Text: *See response to reviewer A above. In addition, we added the following into section 5 (conclusion):* First, we identified substantial heterogeneity in terminology around hospice services between health systems, making it difficult to make inferences about research conducted between different health systems. This finding highlights a need for resources to help scholars and health service users across settings understand similarities and differences across models of care in different health systems. Many existing studies we identified were conducted in settings where hospice care was analogous to palliative care more generally, rather than settings with a distinct hospice model of care different from general palliative care services.

Comment: In line 126 you report that this is the first review of studies done in the hospice setting, whilst others were done in palliative care settings. It is important to note that in some parts of the world these terms are used interchangeably (including in my cited paper) so therefore it cannot be differentiated as being separate entities. If you would like to continue to report that they are different, how are you sure that the papers you have cited as being different to yours are actually different to yours in terms of population studied. Given that there are already many systematic reviews published on a similar topic I wonder if you can say in line 389 that this is the first to examine the landscape of behavioural health research in the hospice setting given that there have been other reviews of the same topic.

Response: We agree with the comments of the reviewer and believe that we address the conceptual issues raised in this comment in the changes described in the comment above and to Reviewer A. In addition, we changed the text to remove any reference to this being the first such study.

Changes in Text*: We removed all references to this being the first such study in the text. We now simply state in the discussion that* "Our study represents a contribution to…" *and that* "We present the findings of…"

Reviewer C:

Comment: Methods: Please outline the steps (Arksey and O'Malley) scoping review, so it is clear that those steps were followed and justify any steps that were different. Also clearly identify whether this study reports on both interrelated reviews, or this reports on just the behavioral health interventions.

Response: We appreciate the reviewer's pointing out that we did not adequately describe the Askey and O'Malley framework. We have added further clarification in the beginning of our methods section (manuscript section 2).

Changes in Text: *In the first paragraph of section 2, we added the following:* Arksey and O'Malley outline a six-step process by which to conduct scooping reviews. (Step 1) The research question was identified by the senior author (a clinician-investigator with expertise in behavioral health and hospice and palliative medicine) in collaboration with the study team and research librarian (see manuscript section 1.3). (Step 2) A research librarian (co-author DW) refined the search strategy based on the research question (see manuscript section 2.1 and Appendix 1). (Step 3) Study selection occurred based on study protocol (see manuscript sections 2.2-2.3). (Step 4) Data were extracted and charted (see manuscript section 2.3). (Step 5) Data were synthesized and presented (see manuscript sections 2.3 and 3). (Step 6) We also utilized the optional expert consultation step noted above.

Comment: Results: In the section on magnitude of change, was there a clinically meaningful change? More context about what the changes meant would be helpful. Also, were there important differences between pharmacologic versus other intervention types? The results might be more meaningful if organized according to intervention type. This might help the reader have a clearer idea of the implications of this review.

Response: We appreciate the reviewer's comments. We agree that providing minimal important differences for the most used scales is important for readers and have added this into the manuscript. Additionally, we have addended the discussion section to better capture the types of interventions that resulted in significant improvements in behavioral health symptoms.

Changes in Text: *Section 3.4, paragraph 2, now reads:* Among those studies demonstrating significant results, four significantly improved only anxiety symptoms (45,54,70,77). Two of these studies were CAM interventions, and two were psychotherapy interventions. Two studies significantly improved only depression symptoms (58,85): one CAM study and one psychotherapy intervention. Six studies significantly improved both anxiety and depression symptoms (56,64,71–73,78). Interventions that improved both anxiety and depression symptoms were CAM (3), spiritual care/other (1), combined psychotherapy and CAM (1), and psychotherapy 1. The remaining five studies improved mood, distress, or overall clinical impression (55,61,63,68,74). These studies consisted of CAM (3), pharmacologic (1), and psychotherapy (1) interventions. Overall, Among the 12 studies that significantly improved depression and/or anxiety symptoms, four employed psychotherapy interventions (45,54,78,85), six employed complementary and alternative medicine interventions (56,58,64,70,71,77), one employed combined psychotherapy and complementary and alternative intervention (72), and one employed spiritual care (73).

Paragraph 3 now reads: Among studies that provided a magnitude of change, improvements in depression and anxiety symptoms were relatively modest. For instance, among the two studies utilizing the HADS and demonstrating significant improvements in depression and anxiety symptoms, the magnitude of change on the HADS subscales ranged from 2.7 to 4.7 points (56,78). Of note, these values exceed the proposed minimally important difference of 1.5 points on each component of the HADS (derived from patients with cardiovascular and pulmonary disease) (86,87). Similarly, studies utilizing the ESAS and demonstrating significant improvements in depression and anxiety symptom scores ranged from 1.0 to 2.3 points change on a 10-point scale (64,77). These changes overlap and may exceed the minimally significant difference in individual ESAS symptom measures (88). Most studies measured improvements immediately after the intervention or study, and few demonstrated sustained improvement in symptoms over time following the intervention.

Comment: Overall, there was little discussion about how health policy affects delivery of care, including hospice care. What would you need to see change with policy that would affect payment and care delivery? Having some information in the discussion about this would be relevant.

Response: We appreciate the reviewer's suggestion and agree that our discussion should be situated in the policy context. We have added the following paragraph to our discussion section.

Changes in Text: *In section 5 (discussion), added:*

Hospice payment and accountability policies in the United States present several barriers to integrating specialist-delivered behavioral health interventions. Hospice care in the United States is paid on a capitated basis (that is, hospice organizations are paid per patient per day of care). Payment for the most common form of hospice care, routine home care, hovers between 150-250 U.S. dollars per patient per day, depending on the duration of the patient's hospice enrollment and the hospice's adherence to quality reporting guidelines (97). Payments delivered to hospice programs must support a range of requisite services provided by an interdisciplinary team, including a physician, a registered nurse, a social worker, and a spiritual care provider (98). In addition to the inherent challenges that the capitated model presents in providing additional behavioral health services, most hospices operating in the United States function in a for-profit model that disincentivizes delivery of non-requisite services (95,99). Hospices participating in quality reporting to the Centers for Medicare & Medicaid Services (CMS) are eligible for higher payment rates. However, hospice quality measures and administrative data do not include behavioral health measures. The consumer survey that constitutes the third quality reporting domain contains 2-4 items about anxiety, sadness, and restlessness/agitation (100). In sum, the care delivery structures in the United States hospice system and payment, quality, and accountability policies may disincentivize research and care delivery contingent on behavioral health specialists. In contrast, some complementary and alternative interventions studied, such as aromatherapy or therapeutic music listening, can be delivered using the existing hospice workforce.

Reviewer D:

Comment: This article is well written but the parts before the intervention section could be re-revised and shortened. If it is part of your cultural context to include research process details, I think only minor changes are needed. Please see my highlighted comments for other changes.

Response: Thank you so much for highlighting this point. We agree that the article is dense, however, we are adhering to the scoping review template of the journal and so have not cut content to adhere to the editorial guidelines.

Changes in Text: N/A

Comment: However, because I think the first several pages talk about process, this material could be shortened considerably. While it is good to know the full research process, it might not need to be included. Personally, I liked the detail, but your message could be lost if there is too much detail included. You will notice that I didn't highlight much toward the intervention section forward of the article. Of course, my comments are my own but can be considered with the rest of the reviewers. If you hear the same comments from my fellow reviewers, this could be a signal to you to make changes to the highlighted sections of this article. Please pass along to your team about the excellent work that you and your team have done on this article. You worked hard on this article and it shows. Because I don't like to stymie hard work, I would just say humbly, "maybe you don't have to work so hard."

Response: See above. We are appreciative of this suggestion and have tried to line-edit the manuscript without cutting any specific, editorially required content. As such, there are a number of minor editorial changes throughout the article reflecting both the reviewer's editorial suggestions sent in the attachment and our own editing.

Changes in Text: Line-edits throughout text.