This manuscript presents a proof of concept for using heatmaps to describe aspects of health care delivery for patients who have died of cancer. The manuscript is premised on an assumption that having more information about variations in treatment patterns will lead to an improvement in quality of care. However, the ways in which one would use more complex measurement results to understand alignment with patient preferences or address inequities in care or outcomes are not clear. These and other concerns are outlined below.

Major Comments

1. Introduction, 2nd paragraph. The authors mention that a limitation of unidimensional measures of end-of-life treatment intensity are that they are based on a reference frame of the health system. However, all of the metrics used to inform the heatmaps are based on health care use metrics — it’s not clear to me that these advance our understanding of misalignment with patient preferences or sources of inequities.

2. More context about the degree to which dynamic heatmaps have been used elsewhere in health care would be useful. This manuscript describes the process of creating heatmaps but would make a greater contribution if it included data on their usability or acceptability, or additional information on how decision makers might act upon these visualizations or identify targets for quality improvement.

3. The meaning of high vs medium vs low hospice, inpatient, and other setting signals is a measure of hospital resources spent on a patient. Without more specific data on patient condition or preferences, it’s not clear to me that these can be mapped to “best” vs “poor” care, as on page 10.

4. It’s unclear whether signals are scaled by patient population, staff availability, or other resource availability metrics.

5. Page 14, line 46. There’s already awareness that costs can vary for similar delivery patterns (and vice versa) — how specifically does a heat map advance our understanding of this mismatch?

6. Without a better understanding of causal relationships among factors included in a heatmap, the way one might act upon the data visualizations isn’t clear to me. It seems highly possible that
the problems of reverse causality, simultaneous determination, and omitted variable bias that we see in more traditional methods would be exacerbated here.

7. Page 17, line 6. The authors suggest that patients would be able to use heatmaps to compare their own utilization to patients similar to them – data or literature to suggest that this is likely/feasible would be useful to include here.

8. In the STROBE checklist, the authors note that sensitivity analyses were not applicable. It seems that different ways of grouping the data or assigning values to groups may change reactions to a heatmap.

Minor Comment
9. The rationale for using data on decedents from a partial year was unclear.

| REVIEWER         | Bentrem, David          |
|------------------|-------------------------|
| NORTHWESTERN UNIVERSITY   |
| REVIEW RETURNED | 09-Oct-2021             |

| GENERAL COMMENTS |
|------------------|
| Khayal et al in their manuscript entitled "Development of dynamic health care delivery heatmaps for end-of-life cancer care: A cohort study" present a model of health care utilization in the last six months of life using systems theory. This work is logical extension of group-based trajectory modeling and a valuable contribution to the literature. |

| REVIEWER         | Samuriwo, Ray           |
|------------------|-------------------------|
| UNIVERSITY OF BRADFORD, SCHOOL OF NURSING AND HEALTHCARE LEADERSHIP  |
| REVIEW RETURNED | 15-Oct-2021             |

| GENERAL COMMENTS |
|------------------|
| General comments Thank you for submitting this article which sets out the novel application of a heat map image depiction system predicated on systems theory to model healthcare utilisation in the last six months of life. This paper made for interesting reading, but certain aspects that need to be addressed or amended as set out in my detailed feedback below, i.e.:

Article Summary
Page 5: There is a very bold statement that ‘Quality improvement decision-makers are likely to find more actionable insights into hospital level care delivery patterns in healthcare delivery heatmaps over conventional quality measures’. In my view this statement is likely to be contested as there are other sources of actionable data that can be used to underpin improvement efforts. Therefore, it may be prudent to use a more measured tone in relation to this statement for example by using the word may instead of the phrase are likely.

Methods
Page 6: The text which reads ‘this decedent research is not human subjects research’ needs to be rephrased so that the point that is being made is set out more clearly

Page 6: Please state the number of people in the 2015-2016 Centers for Medicare and Medicaid Services (CMS) files so that the exact size of the sample is clear. |
Page 6: There needs to be more detail provided to explain what information obtained from each of the four types of CMS files. This is important because the international readership of this journal are unlikely to be familiar with the nature and content of these files.

Patient and Public Involvement
Page 6: There needs to be a more comprehensive rationale provided to explain why there was no patient and public involvement in the design and conduct of this study. It is not sufficient to simply state that there was no patient and public involvement in this study. Patient and public involvement is important in a study like because of its focus. Patients and the public are likely to have a strong interest in studies like this which generate insights that can be used to underpin efforts to improve end of life care especially in the USA where there is strong public interest in Medicare.

Cohort Definition
Page 6: The first two sentences in this section need to be revised so that there is a bit more detail provided about the data and cohort development as well as the previously used methods that were used in this regard. I note the references 11 and 13 refer to the same paper, the only difference being that the former does not have a volume and page number like the latter. Therefore, only one of these two citations should be included in this work and the reference list needs to be updated.

Discussion
Page 16: There is a statement about how heatmaps need not rely on the decedent follow back. This is a bold statement that needs to be discussed in more detail with the analysis and evaluation of a wider range of literature to support the points being made. I also note that one of the claims that is made about the potential wider application of heatmaps refers to summing up patients across any patient cohort group including in relation to race. This is an assertion that needs to be explored in more detail and supported with wider evidence in a more comprehensive manner that is currently the case in this submission. This is important given that is stated that there was no adjustment for race in this study as it is a social construct that may reflect a complex array of epidemiological risk factors such as exposure to racism, economic privation etc. Therefore, there needs to be a more compelling and convincing narrative about how heatmaps can be used more widely for patient cohorts including in relation to race, when that was not the case in this study. In sum, there needs to be a more detailed discussion incorporating a wider range of literature to support the statements that are made about heatmaps and how they could potentially be used more widely. A more detailed discussing would help to establish the original contribution of this study and its potential wider impact with greater clarity.

REVIEWER
May, Peter
Trinity College Dublin, Centre for Health Policy & Management

REVIEW RETURNED
09-Nov-2021

GENERAL COMMENTS
Thank you for this opportunity to review this paper building activity-based heatmaps to characterise end-of-life care in poor-prognosis cancer.
This is a distinctive and innovative approach that seems certain to generate a significant amount of interest as we move further into the era of big data and data visualisation will become more and more valued.

My overarching comment for the paper is to urge the authors to make it as accessible as possible to people unfamiliar with these methods (of whom I am certainly one). An emphasis on plain English and very clear descriptions of what is being done and what it means.

Here are some examples of where clarity might be improved.

ABSTRACT AND ARTICLE SUMMARY

Reading these I was quite confused as to how the individual-level and the hospital-level relate to each other. For example, your objective starts by talking about hospital-level comparisons and then states:

To develop a novel longitudinal, multidimensional visual map of end-of-life care trajectories.

But your Participants are then described individuals. In the design and the results I was not clear if you were talking about individual-level trajectories or hospitals (and if hospitals, how the individual-level data had been aggregated to hospital level). Also unclear was the meaning of longitudinal in this context: for an individual, I assume this to be intensity of care over L6MOL; for a hospital, it could be the L6MOL aggregated for lots of individuals, or it could be how intensity of hospital care for EOL patients is changing at each hospital over time.

A lot of this becomes clear in the paper itself but I didn't find the abstract a clear or compelling description of what was done.

METHODS

The use of the word capability is curious here. As a reader, I think of capability as “the power or ability to do something”.

But in this paper (page 9 or 36, line 28), it is instead the activity of a healthcare system resource for a patient in the form of a subject+verb+operand to describe what action was performed (activity) to whom (patient) and by whom or what (resource).

In other words, it’s about something that has already happened.

Possibly this naming convention pre-dates this paper but it’s such a central part of the whole analysis I found it unhelpful to be so counterintuitively named. E.g. on the next page (line 17) you say “we transformed administrative claims into a set of system capabilities and the times they occurred”.

This sounds like you mapped the capacity of hospitals to provide different types of care. But actually you are describing the
observed provision of different types of care to people. This issue recurs through RESULTS.

DISCUSSION

In limitations there is no mention of the fact that you don’t know individual-level preferences, or how these vary across hospitals, and how variation in preferences may explain (at least some part of) the variation in intensity that you are reporting.

Isn’t it also likely that individual-level events and so need for care varies across hospitals in ways that are not controlled for?

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Dr. Melissa Garrido, James J Peters VA Medical Center, Icahn School of Medicine at Mount Sinai

Comments to the Author:
This manuscript presents a proof of concept for using heatmaps to describe aspects of health care delivery for patients who have died of cancer. The manuscript is premised on an assumption that having more information about variations in treatment patterns will lead to an improvement in quality of care. However, the ways in which one would use more complex measurement results to understand alignment with patient preferences or address inequities in care or outcomes are not clear. These and other concerns are outlined below.

Major Comments

1. Introduction, 2nd paragraph. The authors mention that a limitation of unidimensional measures of end-of-life treatment intensity are that they are based on a reference frame of the health system. However, all of the metrics used to inform the heatmaps are based on health care use metrics — it’s not clear to me that these advance our understanding of misalignment with patient preferences or sources of inequities.

We acknowledge the need to further clarify the use of the term reference frame and its relation to patient preferences or sources of inequities. Our intention was not to juxtapose the reference frame of the healthcare system to patient preferences, but instead to the reference frame of the patient’s trajectory through the disconnected healthcare system. Currently, unidimensional measures typically describe the patient’s use of a specific place/service, such as the intensive care unit (ICU). But to improve quality typically requires making decision changes upstream of a specific location. Consequently, the reference frame of the patient’s trajectory through the different parts of the healthcare system may be useful for a clearer understanding of where to potentially intervene. For example, increasing hospice use can be achieved when non-hospice providers upstream refer patients.

We have re-written the second paragraph to now state:

“Variations in end-of-life treatment intensity have been described using unidimensional measures. Administrative claims data are used to calculate healthcare utilization measures, such as total spending, hospital days, or fraction of patients receiving a particular service (e.g., intensive care unit admission or hospice enrollment). In many cases, measures are not directly actionable (e.g., spending, hospital days). In addition, measures often describe use of a specific healthcare service (e.g., hospital care, hospice care). Consequently, these
measures do not capture the full dimensionality of complex clinical care, where a multidimensional set of providers from multiple specialties provide care across many settings akin to an archipelago of disconnected healthcare islands. Consequently, these unidimensional variables are typically described from the reference frame of a specific healthcare island, yet, improving its measures may require an upstream decision change at a prior island along the patient’s journey from island to island. For example, when increasing hospice utilization is a goal, this can only be achieved through increasing upstream referrals from non-hospice providers. Consequently, understanding a typical upstream utilization pattern can serve as an actionable target for interventions. In other words, improving care quality becomes more actionable with a shift in reference frame from the entities comprising the healthcare system to a patient’s trajectory through the healthcare system. By introducing the element of time, it becomes clearer how services are utilized relative to each other.” [pg. 4]

2. More context about the degree to which dynamic heatmaps have been used elsewhere in health care would be useful. This manuscript describes the process of creating heatmaps but would make a greater contribution if it included data on their usability or acceptability, or additional information on how decision makers might act upon these visualizations or identify targets for quality improvement.

While we are currently investigating the usability or acceptability of dynamic heatmaps for quality improvement, the field of industrial engineering and operations research (IEOR) applied to healthcare has extensively published dynamic information, such as maps and simulations, to inform decision-making. Our manuscript is unique in that we develop a larger system-level view of care delivery using administrative claims data, whereas IEOR typically develops a clinic-level view of acute care delivery using local-hospital data.

1. Zhang X. Application of discrete event simulation in health care: a systematic review. BMC health services research, 18(1):1–11, 2018.

We now address this limitation in the discussion:

“The field of industrial engineering and operations research uses heatmaps and simulations to inform decision making at the level of individual hospital processes. This paper demonstrates the feasibility of extending these approaches by using administrative claims. Future work will be needed to explore the usability and acceptability of these heatmaps by decision makers and to assess the validity of our study team’s valuations of the relative burdensomeness (i.e., location on the y-axis) of particular place-based resources.” [pg. 20]

3. The meaning of high vs medium vs low hospice, inpatient, and other setting signals is a measure of hospital resources spent on a patient. Without more specific data on patient condition or preferences, it’s not clear to me that these can be mapped to “best” vs “poor” care, as on page 10.

It is a fair point that cohort-based measures based on administrative data lack information about patient condition and preferences. Nevertheless, it is a commonly accepted practice to infer that high hospice utilization for longer periods of time for advanced cancer patients is higher quality care than shorter or no hospice utilization [Earle, JCO 2008]. We only applied value terms to the extremes of utilization. As we wrote in the paper “Given modal population preferences for receiving end-of-life
care at home rather than in the hospital, high inpatient signal intensity and low hospice signal intensity would generally correlate with worse end-of-life quality."

We have edited the sentence to more accurately describe “higher” and “lower” quality, instead of “best” and “poor” care and added a reference. The sentence now reads:

“First, inpatient and hospice Low, Med, or High categories were individually assigned a quartile value, where the value of 3 represents “higher” quality, and 1 represents “lower” quality.5” [pg. 11]

4. It’s unclear whether signals are scaled by patient population, staff availability, or other resource availability metrics.

Signals are only scaled to patient population – we did not scale to staff or resource availability. However, we focus on NCI-designated cancer centers to ensure all centers have similar capabilities. We have clarified the second to last paragraph of the dynamic utilization heatmap development section to now read:

“Second, we normalized each hospital-level dynamic utilization heatmap by the number of beneficiaries included in the first summing step.” [pg. 10]

5. Page 14, line 46. There’s already awareness that costs can vary for similar delivery patterns (and vice versa) – how specifically does a heat map advance our understanding of this mismatch?

While costs may vary for different care patterns, several studies have used overall spending as measures of inefficiency (e.g., the use of high end-of-life expenditures as “wasteful” spending). Much of the research incorporating spending implicitly assumes a focus on how much money is spent. One advantage of measuring overall spending is that it provides a single summary that can plausibly describe the intensity of health care delivery. However, a one-dimensional measure such as total spending cannot explain the contributors to spending over time. Our analysis demonstrates that heatmaps provide a visual of how the money is spent over time. Heatmaps provide an accessible view into healthcare utilization within hospitals. For example, heatmaps can identify which hospitals are providing care with a signature of high-cost low-value care versus high-cost high-value care.

To specifically describe how heatmaps advance our understanding of costs, we added the following text in the Results (Case 2) section:

“In other words, while measures of spending focus on how much money is spent, heatmaps provide a visual explanation of how the money is spent.” [pg. 13]

6. Without a better understanding of causal relationships among factors included in a heatmap, the
way one might act upon the data visualizations isn’t clear to me. It seems highly possible that the problems of reverse causality, simultaneous determination, and omitted variable bias that we see in more traditional methods would be exacerbated here.

We agree that these heatmaps are not yet ready for use in organizational quality improvement. Our goal in this paper is to demonstrate feasibility of the calculations and to share our methods. As we note above in response to comment #2, future work will explore usability and impact. The questions about the potential for reverse causality, simultaneous determination, and omitted variable bias are important to address. The value of heatmaps is that they can actually overcome some of these concerns. For example, the temporality of heatmaps may protect against issues relating to reverse causality. Consequently, the heatmaps may now provide insights that were previously unavailable from a series of individual quality measure values. The goal of using a heatmap is to encourage and potentially identify causal relationships, referral patterns, or decision-making that may be explored further.

For example, for the hospital in Figure 1C, the inpatient visit signal (Decide@Inp) shows utilization for the last 90 days of life, with very high intensity starting 12 days before death. Also, the hospice signal (Hospice@Home) increases later than the intensity spike at 12 days for inpatient visits. This suggests that much of hospice referral may be coming from inpatient visits for this hospital. Conversely, the hospital in Figure 1H shows high utilization of hospice and low inpatient utilization. If the viewer was interested in increasing hospice utilization (currently at 60%), the heatmap suggests that this is not likely to happen by increasing referral from inpatient service. Conversely, decision-makers at the hospital in Figure 1C might focus on actions to increase outpatient hospice referrals with a goal of preventing avoidable hospitalizations in the weeks before death.

We have added the following text to the discussion limitation section.

"Dynamic utilization heatmaps do not serve as a prescriptive tool for change. Instead, they serve as a descriptive tool to complement on-the-ground clinical knowledge and provide guidance to explore causal relationships, referral patterns, or decision-making." [pg. 20]

7. Page 17, line 6. The authors suggest that patients would be able to use heatmaps to compare their own utilization to patients similar to them – data or literature to suggest that this is likely/feasible would be useful to include here.

Thank you for this comment. We have added text to give an example of how this is feasible today with references. The additional text states:

"This is feasible today through the CMS Blue Button 2.0 digital technology, which provides Medicare beneficiaries access to their data and allows third-party developers or researchers to develop applications for patients to access and view their historical healthcare utilization.\textsuperscript{39, 41} Future work will explore whether and how patients might use such information." [pg. 18]

8. In the STROBE checklist, the authors note that sensitivity analyses were not applicable. It
seems that different ways of grouping the data or assigning values to groups may change reactions to a heatmap.

This is an interesting point. We constructed the groups taking a formal modeling approach, augmented with data-driven analysis, and finally a consensus-based approach. We started with a formal systems engineering approach which, by definition, identifies the highest-level categories that are mutually exclusive and collectively exhaustive. From there, we removed any categories that in theory may exist but in reality, do not exist in real data (e.g., transport at home). Finally, we incorporated a consensus-based approach that allowed us to decompose categories to elucidate utilization associated with quality measure constructs (e.g., hospice use). For example, we distinguished hospice treatment from home health treatment at home. Because heatmaps include every claim in our dataset and the groups are described at a high level of abstraction, we do not expect that any claim could be categorized into a different group. With regards to assigning values, we do not assign any values to these groups, instead, values are calculated directly from the number of times they appear in the claims data. Consequently, we did not apply any thresholds and the values are continuous. However, this point does raise the question of the ordering of these groups on the y-axis. We ordered the groups (home, outpatient, ED, inpatient) following a general sense of increasing “intensive” care. While the ordering may make the overall heatmap look different, the values contained within are numerically the same.

To clarify the treatment of quantitative variables in sensitivity analyses, we adjusted the STROBE checklist to point to the pages describing this in the Methods section.

Minor Comment
9. The rationale for using data on decedents from a partial year was unclear.
   We included data starting from a partial year and specifically October 2015, because that was the date when the International Classification of Diseases, 10th revision took effect in the US. We revised our text to now state:

   “We identified beneficiaries with poor prognosis cancers who died between April 1, 2016 and December 31, 2016 between the ages 66-99, had continuous inpatient and outpatient Medicare insurance in the last six months of life, and had at least one hospital discharge or at least two clinician visits in the last six months of life with cancer diagnosis codes associated with a high risk of near-term death and at least one hospital admission for cancer care in the last six months of life. We included beneficiaries for whom we had complete 6-month look-back data coded using the International Classification of Diseases, Tenth Revision (ICD10) codes, which began in October 2015. “ [pg. 7]

Reviewer: 2
Dr. David Bentrem, Northwestern University
Comments to the Author:
Khayal et al in their manuscript entitled "Development of dynamic health care delivery heatmaps for end-of-life cancer care: A cohort study" present a model of health care utilization in the last six months of life using systems theory. This work is logical extension of group-based trajectory modeling and a valuable contribution to the literature.
We thank the reviewer for their positive comment.

Reviewer: 3  
Dr. Ray Samuriwo, Cardiff University College of Biomedical and Life Sciences  
Comments to the Author:  
General comments  
Thank you for submitting this article which sets out the novel application of a heat map image depiction system predicated on systems theory to model healthcare utilisation in the last six months of life. This paper made for interesting reading, but certain aspects that need to be addressed or amended as set out in my detailed feedback below, i.e.:  

Article Summary  
Page 5: There is a very bold statement that ‘Quality improvement decision-makers are likely to find more actionable insights into hospital level care delivery patterns in healthcare delivery heatmaps over conventional quality measures’. In my view this statement is likely to be contested as there are other sources of actionable data that can be used to underpin improvement efforts. Therefore, it may be prudent to use a more measured tone in relation to this statement for example by using the word may instead of the phrase are likely.  
   We have revised the text per the reviewer’s suggestion, changing the word likely to “may”. [pg. 3]  

Methods  
Page 6: The text which reads ‘this decedent research is not human subjects research’ needs to be rephrased so that the point that is being made is set out more clearly.  
   To be more clear, we have rephrased the sentence to now read:  
   “The Dartmouth-Hitchcock Health Human Research Protection Program (IRB) determined that this research is not human subjects research because the data is from decedents and not living humans.” [pgs. 6, 21]  

Page 6: Please state the number of people in the 2015-2016 Centers for Medicare and Medicaid Services (CMS) files so that the exact size of the sample is clear.  
   We listed the exact size of the sample in our results section under the beneficiary and hospital statistics. Our text reads: “A total of 10,119 Medicare decedents with poor prognosis cancers were attributed to 54 hospitals, with an average of 187 beneficiaries (SD 124) per hospital and a range of 44 to 633 beneficiaries.” [pg. 12]  

Page 6: There needs to be more detail provided to explain what information obtained from each of
the four types of CMS files. This is important because the international readership of this journal are unlikely to be familiar with the nature and content of these files.

Thank you for this comment. We expanded the data paragraph to include what type of information was obtained from each file. We also point to the Appendix which further details the content of these files. We added the following text:

“From the Master Beneficiary file, we obtained beneficiary-level information including date of death. From the MEDPAR, Carrier, Outpatient, and Hospice files, we obtained dates of service, hospital provider, service provided, and place of service information. Furthermore, Appendix A details the fields that correspond to this information.” [pg. 6, 7]

Patient and Public Involvement
Page 6: There needs to be a more comprehensive rationale provided to explain why there was no patient and public involvement in the design and conduct of this study. It is not sufficient to simply state that there was no patient and public involvement in this study. Patient and public involvement is important in a study like because of its focus. Patients and the public are likely to have a strong interest in studies like this which generate insights that can be used to underpin efforts to improve end of life care especially in the USA where there is strong public interest in Medicare.

We agree that patients are likely to have a strong interest in this research. The purpose of this work was to demonstrate the method. The Data Use Agreement between Dartmouth and the Center for Medicaid and Medicare Services (CMS) significantly limits our ability to involve patients and the public specifically in this initial step of the work. Future work will involve patients and the public in assessing usability of these heatmaps. We clarified the Patient and public involvement section. The text now states:

“The current phase of the work is computational, with a focus on feasibility. The Data Use Agreement with the Center for Medicare and Medicaid Services does not allow us to involve patients and the public at this stage. The next stage of the work will involve patients and the public to provide input on key assumptions and to assess the usability of these heatmaps as quality improvement tools.” [pg. 6]

Cohort Definition
Page 6: The first two sentences in this section need to be revised so that there is a bit more detail provided about the data and cohort development as well as the previously used methods that were used in this regard. I note the references 11 and 13 refer to the same paper, the only difference being that the former does not have a volume and page number like the latter. Therefore, only one of these two citations should be included in this work and the reference list needs to be updated.

We revised our Cohort Definition section and removed the redundant reference. Our expanded text now reads:

“We identified beneficiaries with poor prognosis cancers who died between April 1, 2016 and December 31, 2016 between the ages 66-99, had continuous inpatient and outpatient Medicare insurance in the last six months of life, and had at least one hospital discharge or at least two clinician visits in the last six months of life with cancer diagnosis codes associated
with a high risk of near-term death and at least one hospital admission for cancer care in the last six months of life. We included beneficiaries for whom we had complete 6-month look-back data coded using the International Classification of Diseases, Tenth Revision (ICD10) codes, which began in October 2015. The look-back period was used to identify the time period for which all healthcare utilization data would be captured and comorbidities for risk-adjustment in the parent project.10 We identified patients with poor prognosis cancers to specify a decedent cohort for whom death was more likely attributable to cancer and clinicians would have been aware that time could be short.11 Poor prognosis cancers were defined based on methods from Iezzoni and colleagues10 that were adapted to the current International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM).12 More detailed cohort development is described elsewhere.12” [pg. 7]

Discussion
Page 16: There is a statement about how heatmaps need not rely on the decedent follow back. This is a bold statement that needs to be discussed in more detail with the analysis and evaluation of a wider range of literature to support the points being made. I also note that one of the claims that is made about the potential wider application of heatmaps refers to summing up patients across any patient cohort group including in relation to race. This is an assertion that needs to be explored in more detail and supported with wider evidence in a more comprehensive manner that is currently the case in this submission. This is important given that is stated that there was no adjustment for race in this study as it is a social construct that may reflect a complex array of epidemiological risk factors such as exposure to racism, economic privation etc. Therefore, there needs to be a more compelling and convincing narrative about how heatmaps can be used more widely for patient cohorts including in relation to race, when that was not the case in this study. In sum, there needs to be a more detailed discussion incorporating a wider range of literature to support the statements that are made about heatmaps and how they could potentially be used more widely. A more detailed discussing would help to establish the original contribution of this study and its potential wider impact with greater clarity.

The question about potential impact of these heatmaps is an important one. We have addressed this in our response to Reviewer #1, comment 2. Namely, our goal here is to demonstrate the feasibility of the approach. While we can posit the added value of this approach to measuring quality from a patient’s health care use data, future research is needed to assess their usability and impact of these metrics.

To clarify the issue of race adjustment, this was stated in the section for “End-of-life care quality metrics” and does not apply to the section that follows called “Dynamic utilization heatmap development.” In creating heatmaps, we do not adjust for any factors since we are not trying to produce a single variable that takes other factors into account. Instead, the goal of this work is to elucidate behavioral patterns across a population of patients simply by summing the individual-level utilization heatmaps for patients attributed to a specific hospital. In this paper, we summed all patients attributed to a hospital to produce a single hospital-level heatmap. Just as easily and in a current working paper, we are summing individual-level heatmaps for groups of patients based on their race. This allows us to generate two hospital-level heatmaps, one for white patients and the other for Black patients at the same hospital.

To clarify the issue of decedent follow-back, the point that we sought to make is that our methodology involves aligning time using an index starting point or a terminal ending point. In this paper, we used a terminal death date to create healthcare utilizations that align in time relative to death across patients. In the same way, we could use a start time point, for example, date of diagnosis to produce a 6-month utilization map.
To clarify the wider application of this method, we are discussing the generalizability of the methodology, not the usability or impact of the metrics. We have revised and added the following text within the 3rd paragraph of the discussion:

“Heatmap patterns need not rely on the decedent follow-back method, as they can also be generated based on an index diagnosis. This is because the underlying heatmap methodology generates individual-level heatmaps that align healthcare utilization across patients using either a start or end date. Furthermore, hospital-level heatmaps are simply the sum of aligned individual-level heatmaps. Consequently, several heatmaps could be used to create informative comparative summary information by summing across patients within any patient cohort group, such as age, race, sex, diagnosis, or any provider cohort group (such as physician group practice). [pgs. 17,18]"
dimensionality of complex clinical care trajectories over time that are needed to inform quality improvement efforts. The objective is to develop a novel visual map of end-of-life care trajectories that illustrate multi-dimensional utilization over time.

**Setting:** The United States.

**Participants:** We identified Medicare claims for fee-for-service beneficiaries with poor-prognosis cancers who died between April and December 2016 and received the preponderance of treatment in the last 6-months of life at a National Cancer Institute or National Comprehensive Cancer Network (NCI/NCCN)-designated hospital.

**Design:** For each beneficiary, we transformed each Medicare claim into two elements to generate a two-dimensional individual-level heatmap. On the y-axis, each claim was classified into a categorical description of the service delivered by a healthcare resource. On the x-axis, the date for each claim was converted into the day number prior to death it occurred on. We then summed up individual-level heatmaps of patients attributed to each hospital to generate two-dimensional hospital-level heatmaps. We used four case studies to illustrate the feasibility of interpreting these heatmaps and to shed light on how they might be used to guide value-based, quality improvement initiatives.

**Results:** We identified nine distinct end-of-life care delivery patterns from hospital-level heatmaps based on signal intensity and patterns for inpatient, outpatient, and home-based hospice services. We illustrate that in most cases, heatmaps illustrating patterns of multidimensional healthcare utilization over time provide more information about care trajectories and highlight more heterogeneity than current unidimensional measures.

**Conclusions:** This study illustrates the feasibility of representing multidimensional end-of-life utilization over time as a heatmap. These heatmaps may provide potentially actionable insights into hospital-level care delivery patterns, and the approach may generalize to other serious illness populations.

**ARTICLE SUMMARY:**

Strengths and limitations of this study

- The development of healthcare delivery heatmaps provides detailed multidimensional end-of-life utilization over time.
- Healthcare delivery heatmaps are developed from the same administrative claims data used to calculate quality measures, but offer increased productivity and efficiency.
- Only cancer centers with National Cancer Institute or National Comprehensive Care Network designation were included in this study, therefore findings cannot be generalized to other hospitals providing cancer care. “[pgs. 2, 3]"

**METHODS**

The use of the word capability is curious here. As a reader, I think of capability as "the power or ability to do something". But in this paper (page 9 or 36, line 26), it is instead the activity of a healthcare system resource for a patient in the form of a subject+verb+operand to describe what action was performed (activity) to whom (patient) and by whom or what (resource). In other words, it’s about something that has already happened. Possibly this naming convention pre-dates this paper but it’s such a central part of the whole analysis I found it unhelpful to be so counterintuitively named. E.g. on the next page (line 17) you say "we
transformed administrative claims into a set of system capabilities and the times they occurred”. This sounds like you mapped the capacity of hospitals to provide different types of care. But actually you are describing the observed provision of different types of care to people. This issue recurs through

Thank you for this interesting comment on the use of the term "capabilities". Yes, the term "capabilities" pre-dates this paper and originates in the field of service (delivery) systems where there is a need to identify what activity is done on what and by whom. Healthcare delivery systems are one type of service delivery system. To clarify this definition, we have clarified the text to now state:

“Briefly, a healthcare system capability represents a system’s ability to perform an activity by a healthcare system resource for a patient in the form of a subject+verb+operand to describe what action can be performed (activity) by whom or what (resource) for a patient.” [pg. 9]

It is also important to distinguish the presence of a capability in a system (that exists perpetually) from its occurrence in the form of a completed action at a certain moment in time. Administrative claims data show that certain capabilities occurred and from this data, we can infer the presence of system capabilities in the healthcare delivery system. We avoid the use of the word “capacity” which is most frequently defined in service systems as a quantitative measure of bounding a certain capability. For example, the emergency room has a capacity of 100 patients.

DISCUSSION
In limitations there is no mention of the fact that you don’t know individual-level preferences, or how these vary across hospitals, and how variation in preferences may explain (at least some part of) the variation in intensity that you are reporting. Isn’t it also likely that individual-level events and so need for care varies across hospitals in ways that are not controlled for?

We agree it is worth pointing out how patient preferences can impact hospital-level heatmaps in the discussion. Healthcare utilization in claims simply indicates the decision to utilize service and provides no direct information about the level of shared decision-making between providers and patients and their families or individual-level preferences. The goal of hospital-level heatmaps is not to identify behaviors due to the hospital, but simply to identify the behaviors that emerge for the population attributed to a hospital.

We have included an additional paragraph to address this limitation. Our text now states:

“While dynamic utilization heatmaps elucidate behaviors that emerge from a hospital’s population, these behaviors should not be attributed only to the providers or hospital system. Indeed, healthcare utilization in claims simply indicates the decision to utilize service and provides no direct information about the level of shared decision-making between providers and patients and their families or individual-level preferences. Therefore, when reflecting on behaviors that emerge from these heatmaps, it is essential to question or ascertain decision choices by both the providers/hospital system and patients and their families.” [pg. 20]
VERSION 2 – REVIEW

REVIEWER
Samuriwo, Ray
University of Bradford, School of Nursing and Healthcare Leadership

REVIEW RETURNED
11-Mar-2022

GENERAL COMMENTS
Thank you for submitting this revised paper on the novel application of a heat map image depiction system predicated on systems theory to model healthcare utilisation in the last six months of life. Most of the feedback on the previous submission have been addressed. However, there are two key elements in the methods section and discussion which need to be clarified, i.e.:

Methods
Data section
Line 41-42: There is still no clear statement about the number of people in the 2015-2016 Centers for Medicare and Medicaid Services (CMS) files. It is not sufficient to state that a 100% sample of the 2015-2016 Centers for Medicare and Medicaid Services (CMS) files was used as the international audience of this journal will not be familiar with the sample size. As this is an international journal, the exact sample size needs to be stated in the form of an exact figure.

Discussion
Line 3-33: The points that are made in new dynamic utilization heatmaps must be supported with references. This is important as there are several bold statements that are made about the utilization of heatmaps, with reference to industrial engineering and operations research. These statements about industrial engineering and operations research need a reference so that the reader can evidence underpinning them.

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3

Dr. Ray Samuriwo, University of Bradford

Comments to the Author:
Thank you for submitting this revised paper on the novel application of a heat map image depiction system predicated on systems theory to model healthcare utilisation in the last six months of life. Most of the feedback on the previous submission have been addressed. However, there are two key elements in the methods section and discussion which need to be clarified, i.e.:

Methods
Data section
Line 41-42: There is still no clear statement about the number of people in the 2015-2016 Centers for Medicare and Medicaid Services (CMS) files. It is not sufficient to state that a 100% sample of the 2015-2016 Centers for Medicare and Medicaid Services (CMS) files was used as the international audience of this journal will not be familiar with the sample size. As this is an international journal, the exact sample size needs to be stated in the form of an exact figure.
Thank you for this comment. We have re-written this sentence to explicitly state the sample size:

“We identified Medicare fee-for-service beneficiary healthcare utilization using a 100% sample of 2015-2016 Centers for Medicare and Medicaid Services (CMS) files, which included 126,434 decedents with poor prognosis cancers (defined below).” [pg. 13]

Discussion
Line 3-33: The points that are made in new dynamic utilization heatmaps must be supported with references. This is important as there are several bold statements that are made about the utilization of heatmaps, with reference to industrial engineering and operations research. These statements about industrial engineering and operations research need a reference so that the reader can evidence underpinning them.

Thank you. We have added a total of seven references. The references by Roelens et al. titled “The Creation of Business Architecture Heat Maps to Support Strategy-Aligned Organizational Decisions” and the book by Griffin et al. titled “Healthcare Systems Engineering” directly link the use of heatmaps and industrial and operations methods to support healthcare decision-making. The reference by Lather et al. titled “Framework for a Hybrid Simulation Approach for an Integrated Decision Support System in Healthcare Facilities,” describes the integration of simulation and visualization more broadly for healthcare decision making. The highly referenced annotated bibliography by Klein et al. from 1993 titled “Simulation-modeling and Health-Care Decision Making” highlights a history of more than two decades of simulation modeling to facilitate clinical and operational decision-making—indicating a long history of applying simulation methods to date to support healthcare decision-making. Finally, we included the references by Fakhimi et al. titled “Operations research within UK healthcare: a review” and two classic textbooks titled “Handbook of healthcare operations management” and “Operations research and health care: a handbook of methods and applications” that describe the range of applications that can be applied to healthcare decision-making using simulation and visualization methods.

“The field of industrial engineering and operations research uses heatmaps and simulations to inform decision making at the level of individual hospital processes.” [pg. 20]