



SYSTEMATIC REVIEW

REVISÉD Improving palliative care with machine learning and routine data: a rapid review [version 2; peer review: 2 approved, 1 approved with reservations]

Previously titled: Improving palliative and end-of-life care with machine learning and routine data: a rapid review

Virginia Storick ¹, Aoife O'Herlihy¹, Sarah Abdelhafeez¹, Rakesh Ahmed ¹, Peter May ¹⁻³

¹School of Medicine, Trinity College Dublin, Dublin, D02, Ireland

²Centre for Health Policy and Management, Trinity College Dublin, Dublin, D02, Ireland

³The Irish Longitudinal study on Ageing, Trinity College Dublin, Dublin, D02, Ireland

v2 First published: 15 Jul 2019, 2:13 (<https://doi.org/10.12688/hrbopenres.12923.1>)

Latest published: 12 Aug 2019, 2:13 (<https://doi.org/10.12688/hrbopenres.12923.2>)

Abstract

Introduction: Improving palliative care is a priority worldwide as this population experiences poor outcomes and accounts disproportionately for costs. In clinical practice, physician judgement is the core method of identifying palliative care needs but has important limitations. Machine learning (ML) is a subset of artificial intelligence advancing capacity to identify patterns and make predictions using large datasets. ML has the potential to improve clinical decision-making and policy design, but there has been no systematic assembly of current evidence.

Methods: We conducted a rapid review, searching systematically seven databases from inception to December 31st, 2018: EMBASE, MEDLINE, Cochrane Library, PsycINFO, WOS, SCOPUS and ECONLIT. We included peer-reviewed studies that used ML approaches on routine data to improve palliative care for adults. Our specified outcomes were survival, quality of life (QoL), place of death, costs, and receipt of high-intensity treatment near end of life. We did not search grey literature.

Results: The database search identified 426 citations. We discarded 162 duplicates and screened 264 unique title/abstracts, of which 22 were forwarded for full text review. Three papers were included, 18 papers were excluded and one full text was sought but unobtainable. One paper predicted six-month mortality, one paper predicted 12-month mortality and one paper cross-referenced predicted 12-month mortality with healthcare spending. ML-informed models outperformed logistic regression in predicting mortality where data inputs were relatively strong, but those using only basic administrative data had limited benefit from ML. Identifying poor prognosis does not appear effective in tackling high costs associated with serious illness.

Conclusion: While ML can in principle help to identify those at risk of adverse outcomes and inappropriate treatment, applications to policy and

Open Peer Review

Reviewer Status

	Invited Reviewers		
	1	2	3
REVISÉD version 2 published 12 Aug 2019	 report		 report
	↑		
version 1 published 15 Jul 2019	 report	 report	

- 1 **Tim Luckett** , University of Technology Sydney, Sydney, Australia
- 2 **Wei Gao** , King's College London, London, UK
- 3 **Amara Callistus Nwosu** , University of Liverpool, Liverpool, UK

Any reports and responses or comments on the article can be found at the end of the article.

practice are formative. Future research must not only expand scope to other outcomes and longer timeframes, but also engage with individual preferences and ethical challenges.

Keywords

Machine learning, artificial intelligence, palliative care, terminal care, multimorbidity, quality of life, costs, decision-making



This article is included in the [Ageing Populations](#) collection.

Corresponding author: Peter May (mayp2@tcd.ie)

Author roles: **Storick V:** Conceptualization, Data Curation, Formal Analysis, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **O’Herlihy A:** Conceptualization, Data Curation, Formal Analysis, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Abdelhafeez S:** Conceptualization, Data Curation, Formal Analysis, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **Ahmed R:** Conceptualization, Data Curation, Formal Analysis, Investigation, Writing – Original Draft Preparation, Writing – Review & Editing; **May P:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Resources, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: Health Research Board Ireland [ARPP-A-2018-005 to PM]

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2019 Storick V *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Storick V, O’Herlihy A, Abdelhafeez S *et al.* **Improving palliative care with machine learning and routine data: a rapid review [version 2; peer review: 2 approved, 1 approved with reservations]** HRB Open Research 2019, 2:13 (<https://doi.org/10.12688/hrbopenres.12923.2>)

First published: 15 Jul 2019, 2:13 (<https://doi.org/10.12688/hrbopenres.12923.1>)

REVISED Amendments from Version 1

Minor changes were made to the manuscript between versions 1 and 2, in line with reviewer comments (see responses to reviewers for consideration of each specific change). The most substantial changes were (1) amending “palliative and end-of-life care” to “palliative care” throughout, starting with the title, as a reviewer found the original confusing; (2) a more detailed justification for exclusion of decedent cohort study designs in Methods>Eligibility criteria>Types of studies and reports.

See referee reports

Introduction

Background

Improving care for people with serious and complex medical illness is a health system priority worldwide. Between 2016 and 2060 there will be an estimated 87% increase globally in the number of deaths that occur following serious health-related suffering, with low-income countries experiencing the largest proportional increases¹. Health systems originally configured to provide acute, episodic treatment often provide poor-value care to complex multimorbid populations².

Palliative care is an approach “that improves the quality of life (QOL) of patients and their families facing the problems associated with life-threatening illness, through the prevention and relief of suffering by means of early identification and impeccable assessment and treatment of pain and other problems, physical, psychosocial and spiritual”³. Studies suggest that palliative care improves outcomes and reduces health care costs for people with serious medical illness, although significant gaps in the evidence base persist⁴. Insufficient palliative care capacity is reported even among those nations whose services perform strongly on international rankings^{5,6}, and need will only grow given demographic ageing⁷.

In clinical practice, physician judgement remains the *de facto* method of identifying palliative care needs and predicting adverse outcomes including mortality⁸. However, studies have repeatedly shown that clinicians tend to make imprecise and overly optimistic predictions of survival in metastatic cancer, where prognosis is most accurate of major terminal illnesses⁹⁻¹¹. No subgroup of clinicians are proven to be more accurate than others in late-stage predictions⁸.

For research and policy, primary research studies are rare due to ethical and practical issues, increasing reliance on the use of routine data^{4,12}. Evaluations of programmes and interventions employing traditional analytic approaches encounter challenges in missing data and unobserved confounding¹³.

Health care is entering an era of ‘big data’ in which researchers and providers will have access to unprecedented levels of information on patients¹⁴. Machine learning (ML) is a subset of artificial intelligence that is rapidly advancing capacity to identify patterns and make predictions using large datasets¹⁵. In contrast to traditional analytic methods, where the analyst

specifies data inputs according to hypotheses and/or conceptual models, ML approaches leverage computing power to identify patterns in available data and can make inferences without explicit user instruction¹⁶. These have a well-documented potential to improve clinical decision-making by analysing electronic health records and other routinely collected health data that are commonly large and “dirty”, i.e. erroneous and incomplete¹⁷, although significant data missingness may be still be a source of bias¹⁸.

Rationale and aim

ML approaches to improving decision-making may be strongly appropriate for palliative care. First, patient care is complex with high illness burden and management challenges including polypharmacy¹⁹. ML tools may be effective at identifying patterns in data on complex patients, for example denoting risk of functional decline or mortality, that are not obvious to healthcare professionals with limited time and analytic expertise to interpret available data. Second, the volume of projected palliative care need means that a minority of patients will receive care from palliative care specialists^{7,20}. Tools that can aid non-specialists to identify need and provide appropriate care are essential to avoiding population health crises in an era of demographic ageing²¹. While multiple face-to-face clinical assessment tools exist for this purpose, meeting challenges of scale requires improved use of routinely collected data. Third, in the context of reliance on routine data^{4,12,13}. Since ‘big’ data are largely routine data, optimising analytic approaches to these may be particularly impactful in palliative care compared to fields where randomised trials and large primary data collection are more feasible and commonplace.

Multiple recent editorials in topic journals reflect the growing interest in principle in using big data and ML to improve palliative care²²⁻²⁵. Less has been written on the empirical evidence of these applications. In this context we conducted a rapid review of research studies using ML to improve palliative care. By identifying and organising this evidence for the first time, we anticipated that our findings could inform ongoing and future efforts in this field.

Methods

Eligibility criteria

We included studies meeting the following PICOS (Population, Intervention, Comparison, Outcomes and Study design) criteria.

Types of participants. Studies that reported on adults (≥ 18 years). We specified this criterion as children are a distinct and functionally different population. Studies that reported adults and children separately were eligible, but only results for adults would be considered; studies that pooled adults and children in one sample were excluded²⁶. We sought studies seeking to improve palliative care: care that “improves the quality of life of patients and their families facing the problems associated with a life-threatening illness”³. Studies not addressing palliative care, or evaluating a specific discrete treatment (e.g. stent, chemotherapy) were excluded.

Types of outcomes. We specified three domains of interest: patient outcomes (survival; QoL); caregiver outcomes (survival; QoL); and economic outcomes (costs; receipt of cost-(in)effective treatment, e.g. high-intensity treatment at end of life). We specified these as established measures for quality in palliative care, and so of particular interest to practitioners and policymakers²⁷.

Types of studies and reports. We included the following types of studies: those that used routinely collected data with ML approaches to improve palliative care, connecting our outcomes of interest with patients’ characteristics before or at diagnosis. Any prospective study meeting the other criteria was therefore eligible; retrospective studies were eligible provided that the analysis was conducted counting *forwards*, i.e. defining samples and treatments at baseline. Studies that counted backwards, examining samples defined by characteristics at death or according to outcome, were excluded for three reasons. First, participant selection at death biases substantially derived results by distorting timeframe of analysis²⁸. Second, mortality is an outcome and many characteristics at death are outcomes; they are not independent of treatment choices but instead endogenous and therefore inappropriate as an eligibility criteria for evaluating the treatment²⁹. Third, treatment effect estimates from decedent cohort studies are in practice evidence that “treatment *t* should be provided to population *p*”, where *p* is partly defined by imminent death^{28,29}. They are therefore only useful under scenarios of very good prognostic accuracy, and this is seldom the case in the real world⁸⁻¹¹.

This was a rapid review of published literature. Rapid reviews are a well-established methodology for gathering evidence in a limited timeframe, provided that search methods are transparent and reproducible, clear inclusion criteria are applied, and a rigorous if time-limited appraisal is performed^{30,31}. Only peer-reviewed papers in academic journals were considered eligible. Selection of relevant papers was restricted to English-language publications only. We excluded conference abstracts and book chapters returned by the database search, and we did not search grey literature.

Information sources

Database search. The following electronic databases were searched:

- EMBASE
- MEDLINE
- Cochrane Library
- PsycINFO
- Web of Science (WOS)
- SCOPUS
- ECONLIT

Searches were conducted on February 14th, 2019. All databases were searched from their inception dates to December 31st, 2018.

A full search strategy is provided in [Table 1](#). Search terms for all seven databases are provided as extended data³².

Study selection

Screening of titles and abstracts. Two reviewers (any two authors) independently screened each title and abstract of retrieved citations based on the eligibility criteria. Subsequently, conflicts were resolved on the basis of consensus. The online reviewer tool [Covidence](#) was used to manage the screening and selection process.

Screening of full-text reports. The second phase of screening involved downloading and assessing the full-text papers for all of the citations retained from the first phase of screening and applying the eligibility criteria. Two independent reviewers (any two of VS, AOH, SA, RA) screened each full-text paper.

Hand searching. One author (PM) searched the references of all included papers by hand, applying the same criteria.

Data extraction. We extracted study design features to a standardised template: setting (country, year, health care environment), aim, principle methods, data sources and sample size. These are presented in [Table 2](#).

Data synthesis. Narrative synthesis was employed to review the included studies and combine their key findings. Narrative synthesis involves appraisal of all relevant material, grouping

Table 1. Example search strategy, EMBASE.

#	Limiters
1	'palliative therapy'/exp OR 'terminal care'/exp OR 'terminally ill patient'/exp OR 'hospice'/exp
2	Palliat*:ti,ab
3	('terminal illness' OR 'end of life' OR 'end-of-life' OR 'end-stage disease' OR 'last year of life'):ab,ti
4	#1 OR #2 OR #3
5	'machine learning'/exp
6	('data mining' OR 'artificial intelligence' OR 'machine learning' OR 'deep learning' OR 'neural networks'):ti,ab
7	#5 OR #6
8	#4 AND #7
9	Limiters on #8: To end of 2018; articles and reviews and articles in press only. NOT conference proceedings or a book chapters.

Table 2. Key characteristics of included studies.

Lead author (year)	Setting	Aim	Principle methods	Data and sample	Key Results
Einav (2018) ³³	US; random sample of Medicare fee-for-service beneficiaries in 2008	To analyse healthcare spending by predicted 12-month mortality, i.e. can high end-of-life care costs be identified <i>ex ante</i>	Ensemble of RF, gradient boosting and LASSO	Administrative data: demographics, ICD codes, chronic conditions, prior utilization for a baseline sample of 5,631,168 Trajectories of health care use and diagnosis in the prior 12-month period were included	ML model attributed a higher risk score to those who died within one year than those who did not in 87% of cases End-of-life spending is high but deaths do not account heavily for high spending in Medicare overall Focusing on end-of-life spending is not a useful way to identify inappropriate treatment choices
Makar (2015) ³⁴	US; Medicare fee-for-service beneficiaries in 2010	To quantify six-month mortality risk in four disease cohorts: cancer, COPD, CHF, dementia.	Six ML approaches and logistic regression were used in each cohort, of which RF models performed best in primary analysis.	Administrative data: demographics, ICD codes, chronic conditions, functional status, durable medical equipment, prior utilization for 20,000 randomly selected subjects in each disease cohort Traditional baseline characteristics were augmented with values in prior 12-month period, thus capturing disease progression, functional decline, etc.	ML model attributed a higher risk score to those who died within six months than those who did not in 82% of cases Augmented variables key to predictive power; models using only traditional variables were less accurate
Sahni (2018) ³⁵	Minnesota, US; Six-hospital network (one large tertiary care centre; five community hospitals), 2012–2016	To quantify 1-year mortality risk in a cohort of clinically diverse hospitalized patients.	RF models and logistic regression were applied separately and performance compared.	Electronic medical record data, including vital signs, blood count, metabolic panel, demographics and ICD codes for 59,848 patients	ML model attributed a higher risk score to those who died within one year than those who did not in 86% of cases RF model outperforms logistic regression Demographic and lab data key to predictive power; models using ICD codes alone are less reliable

US: United States; RF: random forest; ML: machine learning; COPD: chronic obstructive pulmonary disease; CHF: congestive heart failure; ICD: international classification of disease; SEER: Surveillance, Epidemiology and End Results.

the findings into a coherent thematic narrative. We chose this approach *ex ante* in the context of our broad aims: to understand how ML has been used to improve palliative care to date, and to consider the implications for future research and practice. At no point did we anticipate quantitative results for combination in meta-analysis, since both available data and ML methods would not be standardised in a way that permitted meaningful pooling.

Results

Database search

The database search identified 426 citations for consideration against the eligibility criteria for this review. Following removal of duplicates (n=162), 264 unique citations were forwarded for title and abstract screening. Of these, 242 citations were excluded based on title and abstract, as they clearly did not meet the review’s pre-specified eligibility criteria. A full-text review of the remaining 22 citations was performed, following

which a further 18 citations were excluded and one³⁶ was unobtainable. We corresponded via email with the author of the unobtainable text, who confirmed she did not have access to a copy of the manuscript and was not able to source one.

Full details for each of the 18 exclusions is provided as extended data³². Reasons for exclusion at full text review were wrong methods used, e.g. not using artificial intelligence/computer learning in the study design (n=6); wrong intervention, e.g. a named drug, a stent, a surgery or chemotherapy (n=5); wrong study design, i.e. defining the sample by outcome (n=4); wrong population, i.e. under 18 (n=2); and not a peer-reviewed article (n=1). Three studies were therefore judged as eligible and included in narrative synthesis³³⁻³⁵; a hand-search of all references in these three papers identified no further studies of interest.

The review process is displayed in a PRISMA flow chart (Figure 1).

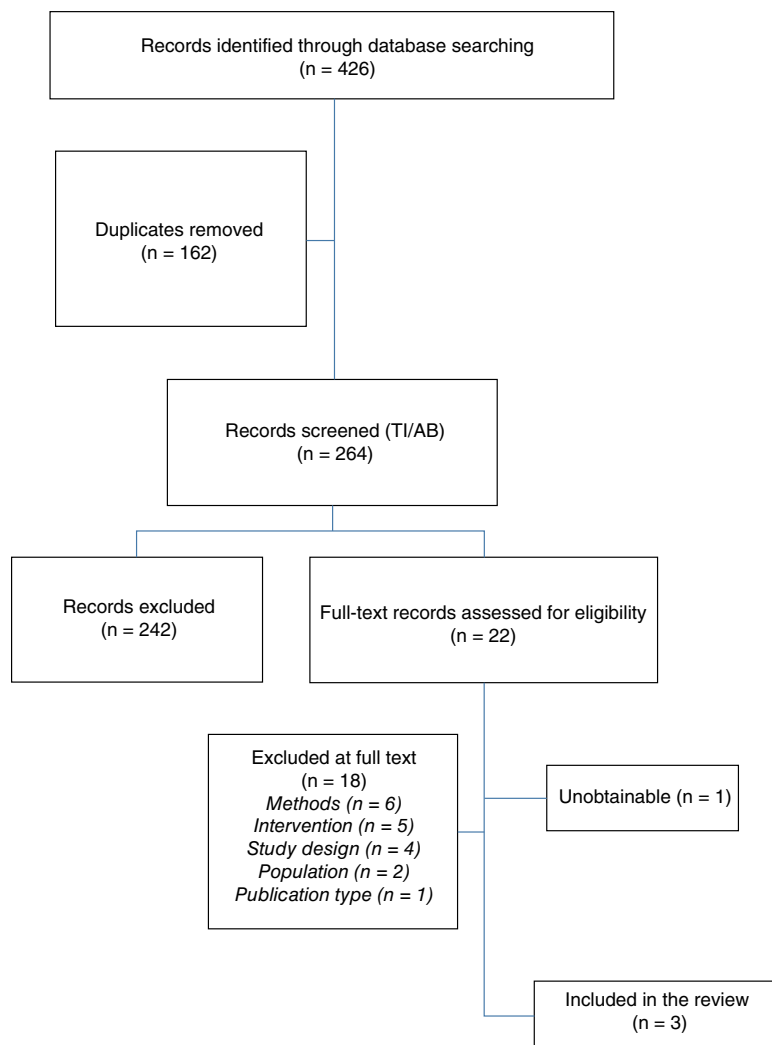


Figure 1. PRISMA flow diagram.

Included studies

Three studies meeting the criteria are presented in [Table 1](#). All came from the United States, two^{33,34} using data from the national Medicare programme for those aged 65 and over, and one³⁵ data on adults admitted to a network of six hospitals in a single urban area. Sampling strategy and sample sizes varied: 59,848³⁵ hospitalised adults (18+); 80,000³⁴ Medicare beneficiaries divided equally into four disease cohorts (cancer, chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), dementia); and a representative sample of 5,631,168³³ Medicare beneficiaries.

ML methods were similar: one³⁵ used random forest (RF) models only, one³⁴ used six approaches in which RF models performed best, and one³³ used an ensemble method combining RF, LASSO and gradient boosting approaches. All three studies used area-under-the-curve (AUC) as a measure of predictive accuracy. In this context AUC represents the probability that a randomly selected patient who died within the time horizon had a higher risk-of-mortality score than one who did not. Two^{34,35} studies compared the performance of their ML models to traditional logistic regression (LR) in predicting mortality.

Main findings

In predicting 12-month mortality amongst a cohort of hospitalised adults, Sahni *et al.* report an AUC of 0.86 with ML methods versus 0.82 with logistic regression³⁵.

In predicting six-month mortality amongst disease cohorts of Medicare beneficiaries, Makar *et al.* report AUC scores of 0.83 (cancer), 0.81 (COPD), 0.76 (CHF) and 0.72 (dementia)³⁴. This performance is superior to LR using the authors' own dataset. Sensitivity analyses suggest that unmeasured disease severity accounts for the difference in model performance across disease cohorts.

In analysing association between predicted 12-month mortality and health care expenditures in a representative sample of Medicare beneficiaries, Einav *et al.* report a weak relationship³³. Their AUC for 12-month mortality is 0.87 but for a population with low mortality rates (about 5% of Medicare beneficiaries died in 2008 and this proportion has since fallen), the 1% of beneficiaries with the highest mortality risk (46%<) account for less than 5% of programme expenditures. Nearly half of this "high-risk" group did not die in the 12-month time horizon. High spending is not concentrated among decedents and *ex ante* identification of mortality does not appear a useful way to identify poor value care.

Variable selection

All papers additionally evaluate variable selection in their models.

Sahni *et al.* use a model in primary analysis that includes as predictors both routinely collected administrative data (demographics; International Classification of Disease (ICD) codes identifying presence of specific conditions) and clinical data (physiologic variables such as blood pressure, respiratory rate,

body mass index; and biochemical variables such as potassium, creatinine, albumin, white blood count, platelet count).

Age, blood urea nitrogen, platelet count, haemoglobin and creatinine were the most important predictors of 12-month mortality. Comorbidities were the least important domain for additional variables. When the model was restricted to comorbidity variables only, predictive power of the model dropped significantly (AUC=0.64)³⁵.

Makar *et al.* use as predictors routinely collected administrative data: demographics, ICD codes, chronic conditions, functional status, durable medical equipment and prior healthcare utilization. In their primary analysis, they not only employ these variables cross-sectionally at baseline but derive 'augmented' versions that capture trajectories, e.g. when diagnoses were made (and so how long the person is living with the condition at baseline), when functional limitations set in, when prior ED visits occurred, etc..

Augmented variables are important for model accuracy. All six models employing ML approaches lose predictive power when relying on traditional baseline variables only, e.g. the RF AUC for cancer drops to 0.77³⁴.

Einav *et al.* employ a similar approach to Makar *et al.*, taking Medicare database variables at baseline and incorporating some prior trajectories. Additionally, they perform a sensitivity analysis with an additional variable of observed outcome, weighted to 10%. While this endogenous predictor improves the AUC to 0.96, it does not improve the model's predictive powers for high Medicare spending³³.

Model choice in the context of available variables

Sahni *et al.* and Makar *et al.* both report superior performance from RF models to traditional LR using their full datasets.

However, as noted above, RF model performance is contingent on available predictors. It is notable that when LR and RF models are compared with fewer available predictors, the superiority of RF approaches is less clear.

For Sahni *et al.*, LR outperforms RF (AUC 0.71 and 0.64 respectively) when the only available clinical data are routinely collected ICD codes³⁵. For Makar *et al.*, LR performs better with only traditional variables (AUC=0.73) than with augmented (AUC=0.56). It also performs much more favourably compared to six ML approaches when only traditional variables are available.

Discussion

Key results

Palliative care is a field of health that may be improved through ML techniques that identify patterns and make predictions using large and complex data. A systematic search of peer-reviewed papers on this topic identifies three relevant studies with important insights.

First, ML approaches are powerful in predicting mortality in older and/or hospitalised adults. All three studies report AUC statistics that compare favourably with prior mortality prediction attempts.

Second, in reported studies these ML methods are superior to traditional LR, but only provided sufficient data are available. Particularly notable were the results of Sahni *et al.*, for whom RF model performance relied on physiologic and biochemical data. When these data were excluded, LR performance was superior to RF and this performance was modest. Many studies have used comorbidity counts derived from ICD codes as predictors of important outcomes for people with serious chronic illness including short-term mortality³⁷⁻³⁹, but the papers in this review suggest these indices may not be as powerful predictors as previously understood.

Makar *et al.* similarly observe much improved RF model performance with more considered inputs, though it is notable that these were not additional clinical data but simply an inventive approach to handling routine data. The predictive power of their four models varied across disease, reflecting unmeasured underlying severity, further emphasising that input data quality decides model performance.

Third, strong predictive power does not correspond to policy implications. Einav *et al.* was the only included paper seeking to apply explicitly its mortality predictions to a policy problem (high costs near end of life), but found a strong mortality prediction model was not useful for their specific purpose.

Corresponding implications for researchers and practitioners seeking to follow these examples are clear. Certainly, ML approaches have significant potential to improve identification of trajectories in seriously-ill populations. However, be mindful of the truism that a model is only as good as its inputs. Where possible, collect granular, individual-level clinical data. Optimise available routine data with innovative approaches where possible. Do not assume that ML approaches are *de facto* superior to traditional LR, and compare performance of different models. Perhaps most important, recognise that improving clinical decision-making will require more than simply improving the predictive power of mortality models. Expand analyses to other outcomes of interest and to the processes underpinning those outcomes.

Limitations

This review has a number of important limitations. We conducted a rapid review of published peer review literature and not a full systematic review incorporating grey literature. Rapid reviews are a well-established methodology for gathering evidence in a limited timeframe³⁰, and we considered this appropriate for our aims of characterising the landscape of a fast-emerging field³¹. Our results can inform ongoing attempts by researchers and practitioners to harness the power of ML methods in improving palliative care, a policy priority worldwide. A particular strength of our methods was that we retained two independent reviewers at each stage, per systematic review

methodology, an approach many rapid reviews eschew⁴⁰. Nevertheless this is a rapidly evolving field^{16,41}, and there are likely conference abstracts and other works in progress that would have featured in a systematic review but are excluded here.

Another limitation of the rapid review timeframe is that we did not formally assess quality or risk of bias in included papers. Instead we incorporated considerations of data quality and study usefulness in reporting, e.g. in interpreting differences in available data across studies in the context of their results. While we specified at the outset that we would not conduct a meta-analysis, we do not consider this a serious limitation in the context of our findings: specific data inputs vary across included papers yet are central to the performance of ML methods; pooling results from these papers is therefore of limited relevance.

Our eligibility criteria led to some exclusions of papers applying ML methods to improve palliative care (see extended data²² for 18 papers excluded at full text review). In particular, four papers that were otherwise-eligible and examined patient outcomes, e.g. QoL, a comfortable death and related factors, were excluded for defining their sample by vital status (a sub-sample of decedents were extracted *ex post*)⁴²⁻⁴⁵. While decedent cohort studies have well established value in some research contexts¹², we established this criterion *ex ante* given endogeneity and bias concerns^{28,29}. This decision was validated by the demonstration of one included study that found analyses of the *ex post* dead are not particularly useful in analysing *ex ante* the sick³³. Additionally we excluded one otherwise-eligible paper for pooling children and adults in the sample⁴⁶, but this decision was based in standard research practice in this field²⁶ and this was reflected in 19 of 21 (90%) full texts using adulthood as an eligibility criterion. Finally, we did not include studies where the endpoints of interest were process measures, e.g. patient-physician interaction, advance care planning and expression of patient/family preferences⁴⁷. While it is rational to assume that improved processes using ML will lead to improved outcomes in the long run^{16,41}, we required that this effect on outcomes be evaluated to be eligible for our review.

One paper was excluded as unobtainable within the timeframe of the analysis, although we did subsequently obtain a copy. We did not consider the paper met our eligibility criteria. While we decided at the outset to exclude papers not in the English language, no paper was ruled ineligible on that basis.

Future research

At the outset of this review we identified outcomes of interest in three domains: patient outcomes (survival; QoL); caregiver outcomes (survival; QoL); and economic outcomes (costs, receipt of cost-(in)effective treatment, high-intensity treatment at end of life). We specified these as established measures for quality in palliative care, and so of direct relevance to practitioners and policymakers²⁷. Most of these domains were unaddressed and so stand as priority areas for future work.

Our review included three studies with a predominant interest in patient mortality. No included study examined patient

outcomes such as QoL. Evaluations of QoL are not straightforward because the outcome of interest is an individual and subjective concept where mortality is an observable binary state. Nevertheless, the reality is that living and dying with serious illness, and caring for those populations, is messy and complex. Studies characterising palliative care need beyond mortality, for example those at risk of pain or unmet need or death anxiety⁴²⁻⁴⁵, and accurately predicting risk would have the capacity to improve clinical decision-making and treatment pathways.

No included study, or any study rejected at full text, incorporated caregiver perspective. This is perhaps not surprising as the patient is naturally the primary focus of clinical practice and associated research. Nevertheless the role of unpaid carers is well recognised in this field, and evaluations of dyad and family outcomes is increasingly common⁴⁸. Identifying caregiver needs in advance would also have vast potential benefit.

One study examined an economic dimension – the long-established association between end-of-life phase and high costs. The authors conclude that this is not a useful lens by which to improve resource allocation in the care of the seriously ill. Rather, those dying with terminal illness and multimorbidity are a subset of all people living with high illness burden and high associated health care use. Identification of appropriate care and supports for this group using ML must be cognisant of recent work identifying substantive treatment effect heterogeneity by disease profile and burden in palliative care⁴⁹. In turn, this review has important insights for treatment effect heterogeneity work that has been based on routinely collected clinical data such as ICD codes: these data, it appears, are relatively weak predictors of relevant outcomes⁵⁰.

An additional concern for future studies across all outcomes of interest is timeframe. The three included studies here used six- and 12-month survival as their outcomes of interest, but there is increasing recognition that palliative care has benefits across the trajectory of life-limiting illness³. For example, American Society of Clinical Oncology guidelines recommend that palliative care is provided across the disease trajectory⁵¹, and the last 12 months of life accounts for less than a third of US cancer care costs⁵². Treatment choices from diagnosis have the greatest scope to impact outcomes and costs^{53,54}, and so studies that can inform these choices are the most useful.

Ethical issues receive scant attention in the studies read for this review. Questions arise with respect to how systems will use the data collected. For example, demographics may be important drivers of outcomes and treatment effects in the context of different experiences and preferences for care across social groups^{55,56}. But offering or providing treatments to some groups and not others based on sociodemographic characteristics is, to say the least, ethically problematic. Also important are the data not collected. Fulfilling the goal of a “good death” involved fixed and modifiable dimensions of patient personality, experience and preferences. Personal resources associated with improved QoL near end of life may include religiosity and beliefs, “acceptance of reality”, “life meaning and purpose”, “self-worth”, “hope”, and “caregivers’ support and acceptance”⁵⁷.

These limitations may be best addressed in the context of broad conceptual considerations for the optimal applications of big data⁵⁸. In all fields it is critical to ground big data collection and measurement in conceptual frameworks to guard against results that are specious, not generalisable or not actionable⁵⁹. This is never more true than in palliative care, where many areas of data collection and conceptual understanding are formative. No matter how powerful the artificial intelligence driving models to improve decision-making, it remains paramount that research, policy and practice protects space for the human wisdom of patients, their families and health care professionals in optimising experience of a unique life event.

Conclusion

Improving palliative care is a policy priority worldwide. ML has the potential to support clinicians in improved decision-making by identifying those at heightened risk of inappropriate care, poor outcomes and mortality. To date studies have demonstrated capacity to improve mortality prediction. Other outcomes have not received equivalent attention. Applications of ML approaches to policy and practice remains formative. Derived results depend on available data and must be interpreted in this context. Future research must not only expand scope to consider other outcomes and longer timeframes, but also address individual needs and preferences in the context of prognosis, and engage with the profound ethical challenges of this emerging field.

Data availability

Underlying data

All data underlying the results are available as part of the article and no additional source data are required.

Extended data

Open Science Framework: Appendix to: [Improving palliative care with machine learning and routine data: a rapid review]. <https://doi.org/10.17605/OSF.IO/AC6TR>

This project contains the following extended data:

- 20190601 Appendix.doc (Summary of search strategies; details of studies rejected at full text review)³²

Reporting guidelines

Open Science Framework: PRISMA checklist and diagram for ‘Improving palliative care with machine learning and routine data: a rapid review’, <https://doi.org/10.17605/OSF.IO/AC6TR>³²

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](https://creativecommons.org/licenses/by/4.0/) (CC0 1.0 Public domain dedication).

Grant information

Health Research Board Ireland [ARPP-A-2018-005 to PM].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Knaul FM, Bhadelia A, Rodriguez NM, *et al.*: **The Lancet Commission on Palliative Care and Pain Relief: findings, recommendations, and future directions.** *Lancet Glob Health.* 2018; 6(S1): S5–S6.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Lehnert T, Heider D, Leicht H, *et al.*: **Review: health care utilization and costs of elderly persons with multiple chronic conditions.** *Med Care Res Rev.* 2011; 68(4): 387–420.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. World Health Organization: **WHO Definition of Palliative Care.** Published 2017. Accessed August 15th, 2017.
[Reference Source](#)
4. Higginson IJ, Evans CJ, Grande G, *et al.*: **Evaluating complex interventions in end of life care: the MORECare statement on good practice generated by a synthesis of transparent expert consultations and systematic reviews.** *BMC Med.* 2013; 11: 111.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Economist Intelligence Unit: **The Quality of Death Index: Ranking palliative care across the world.** 2015.
[Reference Source](#)
6. Arias-Casais N, Garralda E, Rhee J, *et al.*: **EAPC Atlas of Palliative Care in Europe 2019.** Vilvoorde. 2019.
[Reference Source](#)
7. Etkind SN, Bone AE, Gomes B, *et al.*: **How many people will need palliative care in 2040? Past trends, future projections and implications for services.** *BMC Med.* 2017; 15(1): 102.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
8. White N, Reid F, Harris A, *et al.*: **A Systematic Review of Predictions of Survival in Palliative Care: How Accurate Are Clinicians and Who Are the Experts?** *PLoS One.* 2016; 11(8): e0161407.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Glare P, Virik K, Jones M, *et al.*: **A systematic review of physicians' survival predictions in terminally ill cancer patients.** *BMJ.* 2003; 327(7408): 195–198.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Clément-Duchêne C, Carnin C, Guillemin F, *et al.*: **How accurate are physicians in the prediction of patient survival in advanced lung cancer?** *Oncologist.* 2010; 15(7): 782–789.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
11. Gripp S, Moeller S, Bölle E, *et al.*: **Survival prediction in terminally ill cancer patients by clinical estimates, laboratory tests, and self-rated anxiety and depression.** *J Clin Oncol.* 2007; 25(22): 3313–3320.
[PubMed Abstract](#) | [Publisher Full Text](#)
12. Earle CC, Ayanian JZ: **Looking back from death: the value of retrospective studies of end-of-life care.** *J Clin Oncol.* 2006; 24(6): 838–840.
[PubMed Abstract](#) | [Publisher Full Text](#)
13. May P, Garrido MM, Del Fabbro E, *et al.*: **Evaluating Hospital Readmissions for Persons With Serious and Complex Illness: A Competing Risks Approach.** *Med Care Res Rev.* 2019; 1077558718823919.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Raghupathi W, Raghupathi V: **Big data analytics in healthcare: promise and potential.** *Health Inf Sci Syst.* 2014; 2(1): 3.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
15. Arel I, Rose D, Karnowski T: **Deep Machine Learning - A New Frontier in Artificial Intelligence Research [Research Frontier].** 2010; 5(4): 13–18.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Chan A, Chien I, Moseley E, *et al.*: **Deep learning algorithms to identify documentation of serious illness conversations during intensive care unit admissions.** *Palliat Med.* 2019; 33(2): 187–196.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Miotto R, Wang F, Wang S, *et al.*: **Deep learning for healthcare: review, opportunities and challenges.** *Brief Bioinform.* 2018; 19(6): 1236–1246.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Gianfrancesco MA, Tamang S, Yazdany J, *et al.*: **Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data.** *JAMA Intern Med.* 2018; 178(11): 1544–1547.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Institute of Medicine: **Dying in America: improving quality and honoring individual preferences near the end of life.** Washington, DC: The National Academies Press; 2014.
[Reference Source](#)
20. Brereton L, Clark J, Ingleton C, *et al.*: **What do we know about different models of providing palliative care? Findings from a systematic review of reviews.** *Palliat Med.* 2017; 31(9): 781–797.
[PubMed Abstract](#) | [Publisher Full Text](#)
21. Maas EA, Murray SA, Engels Y, *et al.*: **What tools are available to identify patients with palliative care needs in primary care: a systematic literature review and survey of European practice.** *BMJ Support Palliat Care.* 2013; 3(4): 444–451.
[PubMed Abstract](#) | [Publisher Full Text](#)
22. Tanuseputro P: **Delivering care to those in need: Improving palliative care using linked data.** *Palliat Med.* 2017; 31(6): 489–491.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. Nwosu AC, Collins B, Mason S: **Big Data analysis to improve care for people living with serious illness: The potential to use new emerging technology in palliative care.** *Palliat Med.* 2018; 32(1): 164–166.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
24. Davies JM, Gao W, Sleeman KE, *et al.*: **Using routine data to improve palliative and end of life care.** *BMJ Support Palliat Care.* 2016; 6(3): 257–62.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
25. von Gunten CF, Teno JM, Sean Morrison R: **Big Data and End-of-Life Care: Promise and Peril.** *J Palliat Med.* 2016; 19(12): 1240.
[PubMed Abstract](#) | [Publisher Full Text](#)
26. Groh G, Feddersen B, Führer M, *et al.*: **Specialized home palliative care for adults and children: differences and similarities.** *J Palliat Med.* 2014; 17(7): 803–810.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
27. Earle CC, Park ER, Lai B, *et al.*: **Identifying potential indicators of the quality of end-of-life cancer care from administrative data.** *J Clin Oncol.* 2003; 21(6): 1133–1138.
[PubMed Abstract](#) | [Publisher Full Text](#)
28. Bach PB, Schrag D, Begg CB: **Resurrecting treatment histories of dead patients: a study design that should be laid to rest.** *JAMA.* 2004; 292(22): 2765–2770.
[PubMed Abstract](#) | [Publisher Full Text](#)
29. May P, Garrido MM, Cassel JB, *et al.*: **Using Length of Stay to Control for Unobserved Heterogeneity When Estimating Treatment Effect on Hospital Costs with Observational Data: Issues of Reliability, Robustness, and Usefulness.** *Health Serv Res.* 2016; 51(5): 2020–2043.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
30. Ganann R, Cliska D, Thomas H: **Expediting systematic reviews: methods and implications of rapid reviews.** *Implement Sci.* 2010; 5(1): 56.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
31. Khangura S, Konnyu K, Cushman R, *et al.*: **Evidence summaries: the evolution of a rapid review approach.** *Syst Rev.* 2012; 1(1): 10.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
32. May P: **Appendix to: [Improving Palliative and End-of-Life Care with Machine Learning and Routine Data: A Rapid Review].** *OSF.* Web. 2019.
<http://www.doi.org/10.17605/OSF.IO/AC6TR>
33. Einav L, Finkelstein A, Mullainathan S, *et al.*: **Predictive modeling of U.S. health care spending in late life.** *Science.* 2018; 360(6396): 1462–1465.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
34. Makar M, Ghassemi M, Cutler DM, *et al.*: **Short-term Mortality Prediction for Elderly Patients Using Medicare Claims Data.** *Int J Mach Learn Comput.* 2015; 5(3): 192–197.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
35. Sahni N, Simon G, Arora R: **Development and Validation of Machine Learning Models for Prediction of 1-Year Mortality Utilizing Electronic Medical Record Data Available at the End of Hospitalization in Multicondition Patients: a Proof-of-Concept Study.** *J Gen Intern Med.* 2018; 33(6): 921–928.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
36. Tamang S, Kopec D, Shagas G, *et al.*: **Improving palliative care: An information systems approach.** *Journal on Information Technology in Healthcare.* 2005; 3(3): 159–171.
[Reference Source](#)
37. Yurkovich M, Avina-Zubieta JA, Thomas J, *et al.*: **A systematic review identifies valid comorbidity indices derived from administrative health data.** *J Clin Epidemiol.* 2015; 68(1): 3–14.
[PubMed Abstract](#) | [Publisher Full Text](#)
38. Diederichs C, Berger K, Bartels DB: **The measurement of multiple chronic diseases—a systematic review on existing multimorbidity indices.** *J Gerontol A Biol Sci Med Sci.* 2010; 66(3): 301–311.
[PubMed Abstract](#) | [Publisher Full Text](#)
39. Huntley AL, Johnson R, Purdy S, *et al.*: **Measures of multimorbidity and morbidity burden for use in primary care and community settings: a systematic review and guide.** *Ann Fam Med.* 2012; 10(2): 134–141.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
40. Tricco AC, Antony J, Zarin W, *et al.*: **A scoping review of rapid review methods.** *BMC Med.* 2015; 13(1): 224.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
41. Lindvall C, Lilley EJ, Zupanc SN, *et al.*: **Natural Language Processing to Assess End-of-Life Quality Indicators in Cancer Patients Receiving Palliative Surgery.** *J Palliat Med.* 2019; 22(2): 183–187.
[PubMed Abstract](#) | [Publisher Full Text](#)
42. Almasalha F, Xu D, Keenan GM, *et al.*: **Data mining nursing care plans of end-of-life patients: a study to improve healthcare decision making.** *Int J Nurs Knowl.* 2013; 24(1): 15–24.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
43. Lodhi MK, Cheema UI, Stifter J, *et al.*: **Death anxiety in hospitalized end-of-life patients as captured from a structured electronic health record: differences**

- by patient and nurse characteristics. *Res Gerontol Nurs.* 2014; 7(5): 224–234.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
44. Lodhi MK, Ansari R, Yao Y, *et al.*: **Predictive Modeling for Comfortable Death Outcome Using Electronic Health Records.** In: Barbara C, Khan L, eds. *Proc IEEE Int Congr Big Data.* 2015; 2015: 409–415.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
45. Lodhi MK, Stifter J, Yao Y, *et al.*: **Predictive Modeling for End-of-Life Pain Outcome using Electronic Health Records.** In: Perner P, ed. *Adv Data Min.* 2015; 9165: 56–68.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
46. Avati A, Jung K, Harman S, *et al.*: **Improving palliative care with deep learning.** *BMC Med Inform Decis Mak.* 2018; 18(Suppl 4): 122.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
47. Udelsman B, Chien I, Ouchi K, *et al.*: **Needle in a Haystack: Natural Language Processing to Identify Serious Illness.** *J Palliat Med.* 2019; 22(2): 179–182.
[PubMed Abstract](#) | [Publisher Full Text](#)
48. Ornstein KA, Kelley AS, Bollens-Lund E, *et al.*: **A National Profile Of End-Of-Life Caregiving In The United States.** *Health Aff (Millwood).* 2017; 36(7): 1184–1192.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
49. May P, Normand C, Cassel JB, *et al.*: **Economics of Palliative Care for Hospitalized Adults With Serious Illness: A Meta-analysis.** *JAMA Intern Med.* 2018; 178(6): 820–829.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
50. Yourman LC, Lee SJ, Schonberg MA, *et al.*: **Prognostic indices for older adults: a systematic review.** *JAMA.* 2012; 307(2): 182–192.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
51. Ferrell BR, Temel JS, Temin S, *et al.*: **Integration of Palliative Care Into Standard Oncology Care: American Society of Clinical Oncology Clinical Practice Guideline Update.** *J Clin Oncol.* 2017; 35(1): 96–112.
[PubMed Abstract](#) | [Publisher Full Text](#)
52. National Cancer Institute: **Graph by Cancer Site & Phase of Care.** National Institutes of Health. 2018; Accessed December 4th, 2018.
[Reference Source](#)
53. Temel JS, Greer JA, Muzikansky A, *et al.*: **Early palliative care for patients with metastatic non-small-cell lung cancer.** *N Engl J Med.* 2010; 363(8): 733–742.
[PubMed Abstract](#) | [Publisher Full Text](#)
54. May P, Normand C: **Analyzing the Impact of Palliative Care Interventions on Cost of Hospitalization: Practical Guidance for Choice of Dependent Variable.** *J Pain Symptom Manage.* 2016; 52(1): 100–106.
[PubMed Abstract](#) | [Publisher Full Text](#)
55. Frost DW, Cook DJ, Heyland DK, *et al.*: **Patient and healthcare professional factors influencing end-of-life decision-making during critical illness: a systematic review.** *Crit Care Med.* 2011; 39(5): 1174–1189.
[PubMed Abstract](#) | [Publisher Full Text](#)
56. Orlovic M, Smith K, Mossialos E: **Racial and ethnic differences in end-of-life care in the United States: Evidence from the Health and Retirement Study (HRS).** *SSM Popul Health.* 2018; 7: 100331.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
57. Chan WCH: **Can personal resources make a difference? an exploratory study of chinese cancer patients in palliative care.** *J Soc Work End Life Palliat Care.* 2009; 5(3–4): 186–200.
[Publisher Full Text](#)
58. Scott IA: **Hope, hype and harms of Big Data.** *Intern Med J.* 2019; 49(1): 126–129.
[PubMed Abstract](#) | [Publisher Full Text](#)
59. Coveney PV, Dougherty ER, Highfield RR: **Big data need big theory too.** *Philos Trans A Math Phys Eng Sci.* 2016; 374(2080): pii: 20160153.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:   

Version 2

Reviewer Report 15 August 2019

<https://doi.org/10.21956/hrbopenres.14011.r26738>

© 2019 Lockett T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Tim Lockett 

Improving Palliative, Aged and Chronic Care through Clinical Research and Translation (IMPACCT), Faculty of Health, University of Technology Sydney, Sydney, Australia

I am satisfied that the authors have addressed my previous comments.

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Palliative care NOT artificial intelligence.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 14 August 2019

<https://doi.org/10.21956/hrbopenres.14011.r26677>

© 2019 Nwosu A. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Amara Callistus Nwosu 

Palliative Care Institute Liverpool (PCIL), University of Liverpool, Liverpool, UK

The paper is well written, important and timely. The subject of machine learning and the broader topic of technology and artificial intelligence in healthcare is an emerging and important area.

I believe the paper should be indexed and my comments are broad to hopefully help the authors strengthen the paper.

1. The subject area is fairly narrow: Machine learning of routine data in palliative care. Consequently, few articles are included and the authors comment on the articles not included as a potential

limitation. The authors do touch briefly on the reason why only routine data was chosen (i.e. the majority of BigData is routine data); however, it would be useful to have a little more justification for the choice of this (in the background) and a little more about how this may limit the broad usefulness of a paper about 'machine learning' (in the limitations), with some discussion about how analysis of other forms data with machine learning could benefit palliative care (in future research/directions).

2. Implications to policy and practice: I think a headed section of the discussion would be helpful for readers to determine whether there is evidence to change practice/policy. Also, I think it would be useful if the paper could provide a slightly clearer statement of whether the current evidence should change clinical practice or not. Currently it reads that applications to policy and practice are formative. I feel that a clear sentence stating whether clinical practice should change or not based on the evidence would be useful (e.g. 'not enough evidence to recommend changes in policy/practice as evidence is formative etc etc).
3. Future research: It would be useful for the authors to detail the research opportunities of other forms of Big Data (e.g. wearables, apps, user generated data, non health related data (e.g. IoT devices) and other forms of artificial intelligence studies (deep learning, natural language processing).

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Yes

Are sufficient details of the methods and analysis provided to allow replication by others?

Yes

Is the statistical analysis and its interpretation appropriate?

Not applicable

Are the conclusions drawn adequately supported by the results presented in the review?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Applications of technology in palliative care.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Reviewer Report 29 July 2019

<https://doi.org/10.21956/hrbopenres.14002.r26678>

© 2019 Gao W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Wei Gao

Department of Palliative Care, Policy and Rehabilitation, Cicely Saunders Institute, King's College London, London, UK

Major comments

This rapid review provides some interesting findings in a fast-emerging application area in palliative and end of life care. I enjoyed reading it. My major concern is related to the aim of this review which may also affect the choice of the search terms and the study selection.

I think it may be more useful if this is a rapid review on current applications of ML (or even artificial intelligence) in palliative and end of life care. This might simplify the search terms and relax the inclusion criteria, thus may result in more studies (rather than 3 studies only) to be included. It is unclear why the review should be based only on those studies with some specific outcomes.

Minor comments

- Title may need to be changed so it reflects more accurately the aim.
- Type of intervention: care that “improves the quality of life of patients and their families facing the problems associated with a life-threatening illness”. Why “intervention” only? How about observational studies in which case “PECOTS” formula should be used? Why not other outcomes (e.g. satisfaction, functional outcomes, caregiver burden)?
- Types of studies and reports: “Studies that counted backwards, examining samples defined by characteristics at death or according to outcome, were excluded.” It needs more justifications on why these studies were excluded.

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Partly

Are sufficient details of the methods and analysis provided to allow replication by others?

Yes

Is the statistical analysis and its interpretation appropriate?

Yes

Are the conclusions drawn adequately supported by the results presented in the review?

Partly

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Palliative and end of life care; big data; Statistics and Epidemiology; complex intervention.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 06 Aug 2019

Peter May, Trinity College Dublin, Dublin, Ireland

We thank the reviewer for their comments. These have been pasted below, numbered sequentially, and addressed as appropriate with *Response to Reviewer* and/or *Revision to Manuscript*.

We hope that these address the concerns around rationale and conclusions as returned in the original reviewer report.

Major comments

2.1 This rapid review provides some interesting findings in a fast-emerging application area in palliative and end of life care. I enjoyed reading it. My major concern is related to the aim of this review which may also affect the choice of the search terms and the study selection.

I think it may be more useful if this is a rapid review on current applications of ML (or even artificial intelligence) in palliative and end of life care. This might simplify the search terms and relax the inclusion criteria, thus may result in more studies (rather than 3 studies only) to be included. It is unclear why the review should be based only on those studies with some specific outcomes.

Response to Reviewer: As noted in the Methods>Eligibility criteria>Type of outcomes, outcomes were defined broadly to capture those endpoints that palliative care is typically anticipated to impact. We are satisfied that this is choice of outcomes and aim is consistent with research norms in this field. As we note below in response to 2.4, the reviewer interprets incorrectly that some specific outcomes were excluded. We do not consider any relevant outcomes to have been outside our criteria and the supplementary materials show that of 18 papers read and rejected at full text, none was rejected due to choice of outcome. Also, we do not consider it feasible to revisit the aim or terms of the review once it has been completed.

Revision to Manuscript: None.

Minor comments

2.2 Title may need to be changed so it reflects more accurately the aim.

Response: Please see 2.1 response above.

2.3 Type of intervention: care that “improves the quality of life of patients and their families facing the problems associated with a life-threatening illness”. Why “intervention” only? How about observational studies in which case “PECOTS” formula should be used?

Response: Our use of the term ‘intervention’ here is consistent with the PRISMA guidelines and language (see items #4 and #6 in the PRISMA checklist). What the PECOTS framework calls ‘exposure’, PRISMA calls ‘intervention’; in our criteria there is no material difference in use of either term, and there is no related restriction of study design. As noted in the Methods>Eligibility criteria>Type of studies and reports, all comparative study types including observational designs were eligible.

(Please note that in response to Reviewer 1, comment 1.8, Methods>Eligibility criteria>Types of intervention has been merged into the preceding sub-section Types of participants)

Revision: None.

2.4 Why not other outcomes (e.g. satisfaction, functional outcomes, caregiver burden)?

Response: As noted in Methods>Eligibility criteria>Type of outcomes, we specified three domains of interest: patient outcomes; caregiver outcomes; and economic outcomes. Satisfaction and functional outcomes were both eligible outcomes under patient and caregiver outcomes, as sub-categories of quality of life. See also 2.1 response above: no study was excluded due to specifying the “wrong outcomes”.

Revision: None.

2.5 Types of studies and reports: “Studies that counted backwards, examining samples defined by characteristics at death or according to outcome, were excluded.” It needs more justifications on why these studies were excluded.

Revision: Addition to Methods>Eligibility criteria>type of studies and reports:

Studies that counted backwards, examining samples defined by characteristics at death or according to outcome, were excluded for three reasons. First, participant selection at death biases substantially derived results by distorting timeframe of analysis. Second, mortality is an outcome and many characteristics at death are outcomes; they are not independent of treatment choices but instead endogenous and therefore inappropriate as an eligibility criteria for evaluating the treatment. Third, treatment effect estimates from decedent cohort studies are in practice evidence that “treatment t should be provided to population p”, where population p is partly defined by imminent death. They are therefore only useful under scenarios of very good prognostic accuracy, and this is seldom the case in the real world.

Response: As well as the addition to Methods>Eligibility criteria>type of studies and reports, we also note the following lines of justification in the Discussion>Limitations, “While decedent cohort studies have well established value in some EOL research contexts, we established this criterion ex ante given endogeneity and bias concerns. This decision was validated by the demonstration of one included study that found analyses of the ex post dead are not particularly useful in analysing ex ante the sick.”

Competing Interests: No competing interests were disclosed.

Reviewer Report 17 July 2019

<https://doi.org/10.21956/hrbopenres.14002.r26680>

© 2019 Lockett T. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Tim Lockett 

Improving Palliative, Aged and Chronic Care through Clinical Research and Translation (IMPACCT), Faculty of Health, University of Technology Sydney, Sydney, Australia

This manuscript reports a generally well-conducted review that has been reported in accordance with PRISMA guidance. The authors call it a rapid review and seem to justify this by saying that only 'peer reviewed papers in academic journals were considered eligible'. However, I think it would be quite acceptable to call it a systematic review given that many such reviews similarly exclude grey literature on the basis of assured quality.

Regarding terminology, I would use palliative care throughout, as it's confusing for the reader to seemingly distinguish between these without defining the latter.

More specific comments are as follows.

Abstract

- No need to repeat peer-reviewed as an eligibility criterion.
- Poor prognosis is a weak driver of costs.

Introduction

- The final sentence of the first paragraph is a bit harsh – I suggest replacing 'consistently' with 'sometimes'.
- Palliative care is defined as an 'interdisciplinary specialism' but the rest of the definition is concerned with the WHO's palliative approach to care rather than the specialist/generalist nature of the workforce delivering it. I suggest focusing just on the WHO definition.
- I am by no means an expert, but doesn't ML only sometimes 'make inferences without explicit user instruction' - i.e. when it is of an unsupervised kind?

Methods

- I found reference to 'Types of intervention' confusing, as this heading is usually reserved for evaluative reviews (e.g. of RCTs). Given that life-limiting illness was not mentioned under 'Types of participants', I suggest moving this part of the criteria under that heading. QOL is dealt with under 'Types of outcomes'.
- Some justification is needed of excluding studies that 'counted backwards'. This may also need further explanation in the context of the predominance of AUC, which seems to focus on people who died and then look back to see what predicted this.
- The English language requirement needs to be included under eligibility rather than Information sources.

- The Information sources heading needs to also include 'Search'.
- 'Hand searching' is a more widely used term than 'Snowball sampling' for reviews.
- Remove the heading 'Data analysis' and leave in those for extraction and synthesis.

Findings

- Unless I have misunderstood, this section seems to refer to other studies that were not included in the review and so should be moved to the Discussion. But what is meant by 'prior mortality prediction efforts' and which authors are referred to in 'the authors' own dataset'? If the authors of the paper in question, then it should be written in the past tense.
- I lack expertise to comment on interpretations brought to the different statistical approaches used in the studies.

Discussion

- Here, 'prior mortality prediction attempts' really is presumably referring to other research that does not use ML; references are needed.
- I don't quite follow the statement 'This places a question mark against the assumption in much palliative care literature that comorbidity counts derived from ICD codes are important predictors of short-term mortality', and again references are needed.
- The justification that quality appraisal was outside the authors' 'sphere of interest' reads oddly and is not a normally accepted justification. Could it be, instead, that existing quality appraisal tools are not a good fit for studies using routine data?
- I recommend removing reference to meta-analysis in the limitations, as this is totally unsuited to the review's aims.
- I wonder if reference to some other of the issues raised in the following article might be pertinent to the Discussion¹.

References

1. Scott IA: Hope, hype and harms of Big Data. *Intern Med J.* 2019; **49** (1): 126-129 [PubMed Abstract](#) | [Publisher Full Text](#)

Are the rationale for, and objectives of, the Systematic Review clearly stated?

Yes

Are sufficient details of the methods and analysis provided to allow replication by others?

Yes

Is the statistical analysis and its interpretation appropriate?

I cannot comment. A qualified statistician is required.

Are the conclusions drawn adequately supported by the results presented in the review?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Palliative care NOT artificial intelligence.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 06 Aug 2019

Peter May, Trinity College Dublin, Dublin, Ireland

We thank the reviewer for their comments. These have been pasted below, numbered sequentially, and addressed as appropriate with *Response to Reviewer* and/or *Revision to Manuscript*.

GENERAL COMMENTS

1.1 This manuscript reports a generally well-conducted review that has been reported in accordance with PRISMA guidance. The authors call it a rapid review and seem to justify this by saying that only 'peer reviewed papers in academic journals were considered eligible'. However, I think it would be quite acceptable to call it a systematic review given that many such reviews similarly exclude grey literature on the basis of assured quality.

Response to Reviewer: We appreciate this comment. There is an obvious incentive to make this change as an SR ranks higher than an RR. However it feels a little intellectually disingenuous to have defined this as an RR ourselves, conducted the whole thing to a tight time frame (circa 6 months) and excised various elements of an SR that I would consider non-negotiable (grey lit, PROSPERO registration). Despite the temptation then, I'm minded to leave as is unless instructed otherwise by editors and reviewers.

Revision to Manuscript: None.

1.2 Regarding terminology, I would use palliative care throughout, as it's confusing for the reader to seemingly distinguish between these without defining the latter.

Revision: 'Palliative and EOL care' as a sphere of interest has been changed throughout to 'palliative care', including the paper title. A small handful of 'end of life' uses remain where appropriate to specific context.

Abstract

1.3 No need to repeat peer-reviewed as an eligibility criterion.

Revision: Final clause of the Methods deleted.

1.4 Poor prognosis is a weak driver of costs.

Revision: Not clear exactly what this comment implies but perhaps it's a poor explanation of the finding.

Revised final two sentences of the Results:

ML-informed models outperformed logistic regression in predicting mortality where data inputs were relatively strong, but models using only routine administrative data had limited benefit from ML methods. Identifying poor prognosis does not appear effective in tackling high costs associated with serious illness.

Introduction

1.5 The final sentence of the first paragraph is a bit harsh – I suggest replacing ‘consistently’ with ‘sometimes’.

Revision: Changed to ‘often’.

1.6 Palliative care is defined as an ‘interdisciplinary specialism’ but the rest of the definition is concerned with the WHO’s palliative approach to care rather than the specialist/generalist nature of the workforce delivering it. I suggest focusing just on the WHO definition.

Revision: Reference to interdisciplinary specialism removed.

1.7 I am by no means an expert, but doesn’t ML only sometimes ‘make inferences without explicit user instruction’ - i.e. when it is of an unsupervised kind?

Revision: **can** make inferences without explicit user instruction.

Methods

1.8 I found reference to ‘Types of intervention’ confusing, as this heading is usually reserved for evaluative reviews (e.g. of RCTs). Given that life-limiting illness was not mentioned under ‘Types of participants’, I suggest moving this part of the criteria under that heading. QOL is dealt with under ‘Types of outcomes’.

Revision: ‘Type of intervention’ section merged into ‘Type of participants’.

1.9 Some justification is needed of excluding studies that ‘counted backwards’. This may also need further explanation in the context of the predominance of AUC, which seems to focus on people who died and then look back to see what predicted this.

Revision: Addition to Methods>Eligibility criteria>type of studies and reports (also in response to Reviewer 2, comment 2.5):

Studies that counted backwards, examining samples defined by characteristics at death or according to outcome, were excluded for three reasons. First, participant selection at death biases substantially derived results by distorting timeframe of analysis. Second, mortality is an outcome and many characteristics at death are outcomes; they are not independent of treatment choices but instead endogenous and therefore inappropriate as an eligibility criteria for evaluating the treatment. Third, treatment effect estimates from decedent cohort studies are in practice evidence that “treatment t should be provided to population p”, where p is partly defined by imminent death.

They are therefore only useful under scenarios of very good prognostic accuracy, and this is seldom the case in the real world.

Response: Studies using AUC can either do prospectively (enrolling people at baseline and then following them until death and then using baseline data to create and validate a mortality-prediction index) or retrospectively (scooping up all data after the fact then using baseline data to create and validate a mortality-prediction index), and included studies do this. No included studies (but a number of excluded full texts in the supplementary materials) defined the sample at death and counted backwards where any baseline is created ex post by the investigator, which is a recipe for bias (<https://www.ncbi.nlm.nih.gov/pubmed/15585737>).

1.10 The English language requirement needs to be included under eligibility rather than Information sources.

Revision: Moved to end of ‘types of studies and reports’.

1.11 The Information sources heading needs to also include ‘Search’

Revision: Added as Table 1.

1.12 ‘Hand searching’ is a more widely used term than ‘Snowball sampling’ for reviews.

Revision: Changed to ‘Hand searching’.

1.13 Remove the heading ‘Data analysis’ and leave in those for extraction and synthesis.#

Revision: Deleted heading.

Findings

1.14 Unless I have misunderstood, this section seems to refer to other studies that were not included in the review and so should be moved to the Discussion. But what is meant by ‘prior mortality prediction efforts’ and which authors are referred to in ‘the authors’ own dataset’? If the authors of the paper in question, then it should be written in the past tense.

Revision: Reference to papers from outside the review deleted in Results (and referenced in Discussion, see 1.16).

1.15 I lack expertise to comment on interpretations brought to the different statistical approaches used in the studies.

Revision: None.

Discussion

1.16 Here, ‘prior mortality prediction attempts’ really is presumably referring to other research that does not use ML; references are needed.

Revision: Relevant systematic review referenced.

1.17 I don't quite follow the statement 'This places a question mark against the assumption in much palliative care literature that comorbidity counts derived from ICD codes are important predictors of short-term mortality', and again references are needed.

Revision: Sentence revised as: Many studies have used comorbidity counts derived from ICD codes as predictors of important outcomes for people with serious chronic illness including short-term mortality (refs), but the papers in this review suggest these indices may not be as powerful predictors as previously understood.

1.18 The justification that quality appraisal was outside the authors' 'sphere of interest' reads oddly and is not a normally accepted justification. Could it be, instead, that existing quality appraisal tools are not a good fit for studies using routine data?

Revision: I've deleted this sentence as it's not needed or useful.

1.19 I recommend removing reference to meta-analysis in the limitations, as this is totally unsuited to the review's aims.

Response: I tend to agree but a colleague who reviewed a draft of this pre-submission was insistent that it was relevant. So I prefer to include to be clear that we considered it and don't think it's relevant (even though some readers may do so).

1.20 I wonder if reference to some other of the issues raised in the following article might be pertinent to the Discussion.

Revision: Final paragraph before the conclusion is extended:
These limitations may be best addressed in the context of broad conceptual considerations for the optimal applications of big data. In all fields it is critical to ground big data collection and measurement in conceptual frameworks to guard against results that are specious, not generalisable or not actionable. This is never more true than in palliative care, where many areas of data collection and conceptual understanding are formative.

Competing Interests: No competing interests were disclosed.

Comments on this article

Version 2

Reader Comment 19 Aug 2019

Louise Farragher, Health Research Board, Dublin, Ireland

Hi there, I just noticed that the updated title is missing the word 'care'. Congratulations on publishing your protocol and I look forward to reading the final product. Best regards, L Farragher

Competing Interests: None

