

RESEARCH ARTICLE

Observational study of agreement between attending and trainee physicians on the surprise question: “Would you be surprised if this patient died in the next 12 months?”

Christopher J. Yarnell^{1,2,3*}, Laura M. Jewell⁴, Alex Astell⁵, Ruxandra Pinto⁶, Luke A. Devine^{2,7}, Michael E. Detsky^{2,3}, James Downar^{8,9}, Roy Ilan¹⁰, Shail Rawal^{7,11}, Natalie Wong^{3,7,12}, John J. You¹³, Rob A. Fowler^{1,3,6,14}

1 Institute of Health Management, Policy, and Evaluation, University of Toronto, Toronto, Canada, **2** Department of Medicine, Sinai Health System, Toronto, Canada, **3** Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Canada, **4** Memorial University of Newfoundland, Discipline of Family Medicine, Happy Valley-Goose Bay, Canada, **5** University of Manitoba Faculty of Medicine, Section of Critical Care Medicine, Manitoba, Canada, **6** Sunnybrook Health Sciences Centre Department of Critical Care, Toronto, Canada, **7** University of Toronto Temerty Faculty of Medicine, Division of General Internal Medicine, Toronto, Canada, **8** The Ottawa Hospital, Ottawa, Canada, **9** University of Ottawa Faculty of Medicine, Division of Palliative Care, Ottawa, Canada, **10** Department of Critical Care Medicine, Rambam Health Care Campus, Technion, Israel Institute of Technology, Haifa, Israel, **11** University Health Network, General Internal Medicine, Toronto, Canada, **12** Departments of General Internal Medicine and Critical Care Medicine, St Michael's Hospital, Toronto, Canada, **13** Division of General Internal and Hospitalist Medicine, Department of Medicine, Trillium Health Partners, Mississauga, Ontario, Canada, **14** Institute for Clinical Evaluative Sciences, Toronto, Canada

* christopher.yarnell@sinaihealth.ca



OPEN ACCESS

Citation: Yarnell CJ, Jewell LM, Astell A, Pinto R, Devine LA, Detsky ME, et al. (2021) Observational study of agreement between attending and trainee physicians on the surprise question: “Would you be surprised if this patient died in the next 12 months?”. PLoS ONE 16(2): e0247571. <https://doi.org/10.1371/journal.pone.0247571>

Editor: Jason Chia-Hsun Hsieh, Chang Gung Memorial Hospital at Linkou, TAIWAN

Received: August 14, 2020

Accepted: February 10, 2021

Published: February 25, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0247571>

Copyright: © 2021 Yarnell et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant identified data are within the manuscript and its [Supporting Information](#) files.

Abstract

Background

Optimal end-of-life care requires identifying patients that are near the end of life. The extent to which attending physicians and trainee physicians agree on the prognoses of their patients is unknown. We investigated agreement between attending and trainee physician on the surprise question: “Would you be surprised if this patient died in the next 12 months?”, a question intended to assess mortality risk and unmet palliative care needs.

Methods

This was a multicentre prospective cohort study of general internal medicine patients at 7 tertiary academic hospitals in Ontario, Canada. General internal medicine attending and senior trainee physician dyads were asked the surprise question for each of the patients for whom they were responsible. Surprise question response agreement was quantified by Cohen's kappa using Bayesian multilevel modeling to account for clustering by physician dyad. Mortality was recorded at 12 months.

Results

Surprise question responses encompassed 546 patients from 30 attending-trainee physician dyads on academic general internal medicine teams at 7 tertiary academic hospitals in

Funding: The study was funded by a Resident Research Grant from the Physician Services Incorporated Foundation (Dr Yarnell, R15-40). Dr Yarnell is also funded by the Canadian Institutes for Health Research CGS-M, the Clinician Investigator Program, and the Eliot Phillipson Clinician Scientist Training Program at the University of Toronto. The opinions, results and conclusions reported in this paper are those of the authors and are independent from the funding sources including the PSI Foundation. No endorsement by any of the funding agencies is intended or should be inferred. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Ontario, Canada. Patients had median age 75 years (IQR 60–85), 260 (48%) were female, and 138 (25%) were dependent for some or all activities of daily living. Trainee and attending physician responses agreed in 406 (75%) patients with adjusted Cohen's kappa of 0.54 (95% credible interval 0.41 to 0.66). Vital status was confirmed for 417 (76%) patients of whom 160 (38% of 417) had died. Using a response of "No" to predict 12-month mortality had positive likelihood ratios of 1.84 (95% CrI 1.55 to 2.22, trainee physicians) and 1.51 (95% CrI 1.30 to 1.72, attending physicians), and negative likelihood ratios of 0.31 (95% CrI 0.17 to 0.48, trainee physicians) and 0.25 (95% CrI 0.10 to 0.46, attending physicians).

Conclusion

Trainee and attending physician responses to the surprise question agreed in 54% of cases after correcting for chance agreement. Physicians had similar discriminative accuracy; both groups had better accuracy predicting which patients would survive as opposed to which patients would die. Different opinions of a patient's prognosis may contribute to confusion for patients and missed opportunities for engagement with palliative care services.

Introduction

Provision of optimal end-of-life care requires reliable identification of patients likely to die in the near future. Unfortunately, few reliable bedside tools are available to identify patients who may benefit from goals-of-care discussions due to near-term risk of death [1,2]. Predictive models based on illness severity are intended for use at the population level and may be misleading when applied to individual patients [1,3–5]. While some patients suffer from terminal diseases with predictable trajectories, many suffer from multiple comorbidities conferring an unpredictable trajectory towards death [6]. Uncertainty in patient prognosis can increase patient distress, impair communication between physicians and patients, and delay appropriate adoption of a palliative approach [7]. Further, the risk of death may not correlate with other important aspects of patient care such as uncontrolled symptoms or psychological distress [8,9]. Clinician estimates of a patient's risk of death are fallible, although accuracy may correlate with clinician experience, and these estimates likely have a large influence on end-of-life care planning [8,10–15].

One way to gain insight into clinician estimates of prognosis is to investigate how prognostic accuracy varies with training by comparing the prognostic estimates of attending and trainee physicians. Studying differences in prognostic estimates between trainee and attending physicians provides a way to investigate the development of prognostic skills, including whether or not accuracy improves with clinical training. Prognostic discordance and covariates associated with discordance may identify an opportunity for education by highlighting the patient characteristics that make prognosis more difficult. Discordance between trainee and attending physician prognoses also has practical implications because trainee physicians play an important role for inpatients at academic health centres, where discordance may cause confusion in clinical plans, mixed messages for patients, and may partially explain observed differences between received and documented patient end-of-life care preferences [16,17]. The prevalence of discordance between attending and trainee physician prognoses is unknown.

The surprise question is a simple tool to assess patient prognosis and screen for patients who may benefit from end-of-life care [18–21]. The surprise question asks "Would you be

surprised if this patient died in the next 12 months?” and a response of “No” is intended to identify a patient with potentially unmet palliative care needs [22]. In the absence of a gold standard for unmet palliative care needs, most evaluation of the surprise question has instead focused on mortality. A “no” response is correlated with increased mortality rates and discriminatory ability is similar to other prognostic indices intended for hospitalized older adults [1,2,23,24]. Hospitalized patients may be at higher risk of unmet palliative care needs and mortality because of the situation leading to admission and potential gaps in the social safety net that may precipitate hospital admission [25–27]. Improving the identification of patients at high risk of dying can improve clinical care through providing more certainty to patients, more clarity for clinicians making recommendations about invasive interventions or investigations, and triggering earlier appropriate activation of palliative care services. Therefore we compared attending and trainee physician responses to the surprise question in a general internal medicine population to assess the extent and prevalence of discordance as well as the prognostic value of the surprise question with respect to mortality in this population.

Methods

Study population

This was a prospective cohort study of general internal medicine inpatients and their corresponding physicians at 7 tertiary and quaternary academic hospitals in Ontario. The hospitals varied in size (median 463 beds, range 256 to 1325). Admissions to the general internal medicine ward at each hospital occurred primarily through the emergency department as opposed to transfers from other hospitals. The majority of general internal medicine patients at each hospital were cared for by the academic teaching teams involved in this study.

Attending physicians were all certified in internal medicine through the Royal College of Physicians and Surgeons, Canada. Trainee physicians were all second (or higher) year residents in a Canadian internal medicine residency program training for the same certification. There is no mandatory requirement for specific palliative care clinical exposure in internal medicine training in Canada, although residents may participate in clinical electives to gain exposure to this area and “care of the dying” is one of the objectives of training [28].

Recruitment occurred at one vanguard site in 2014 (to hone study procedures and case reporting forms) and at the other sites between April 2017 and October 2018. All patients admitted to general internal medicine teaching teams and their corresponding physician teams were eligible for inclusion and identified on the day of survey. There were no exclusions based on patient language or cognitive status.

Surprise question responses

Clinicians were asked “Would you be surprised if this patient died in the next 12 months?” for each of the patients for whom they were responsible, totalling two responses (one trainee physician response and one attending physician response) per patient. Trainee and attending physicians were surveyed independently with no knowledge of each others’ responses. The year of medical school graduation and duration (in days) on the current service was recorded for each respondent.

Patient data collection

Baseline patient data were collected from the hospital chart including demographic data (gender, age), baseline pre-hospital data gleaned from the clinical notes (type of residence, functional status, marital status), medical information (admitting diagnosis, comorbidities,

admission creatinine), length of stay on the day the surprise question was administered and documented cardiopulmonary resuscitation (CPR) status.

12-month patient mortality data collection

At every site, vital status at 12 months was recorded for each patient by reviewing the hospital electronic record or publicly available online death databases and obituaries. At all but one site, if the 12-month vital status could not be determined from the chart or death databases and obituaries, then a research assistant sent a letter to the patient's address on file offering the opportunity to opt out of a follow-up phone call. A research assistant then attempted to contact by phone all patients with undetermined vital status who did not opt out of the follow up phone call. At one site, no attempts to contact patients after discharge were permitted by the research ethics board.

Ethics

This study was approved by the Research Ethics Boards at Sunnybrook Health Sciences Centre, St Michael's Hospital, Sinai Health Systems and University Health Network, Kingston General Hospital, and Hamilton General Hospital. The waiver for initial patient participation was important to ensure a broad enrolment population that did not exclude vulnerable subgroups such as patients with cognitive impairment, interpretation needs, or mental illness for whom the surprise question may have different operating characteristics and for whom there may be a differential risk of death or unmet palliative care needs [29].

Statistical analysis

The primary analysis assessed agreement between attending and trainee physician 12-month surprise question responses using an adjusted Cohen's kappa measure of interrater reliability (chance-corrected agreement) to account for clustering by physician dyad [30,31]. This was calculated from the posterior probabilities of each response using a multinomial Bayesian regression model with random effects allowing for clustering by physician (**Supporting information**) [32]. Modeling the clustering by physician pair was important in order to identify the extent of variability in agreement across physician dyads [33]. The prior distributions were chosen to be minimally informative. Discordance was assessed by the relative risk of attending physicians responding "No" in discordant cases using the posterior probability distributions.

The accuracy of surprise question responses with respect to 12-month mortality was derived from 2-by-2 tables including calculation of positive and negative likelihood ratios. Bayesian multilevel models were also used to assess the prognostic characteristics of the 12-month surprise question with respect to 12-month mortality for attending physicians and trainee physicians accounting for clustering by physician dyad. This also included calculation of sensitivity, specificity, and likelihood ratios.

A further exploratory model assessed for association between kappa value and selected clinically relevant patient factors: age, sex, functional status, CPR status, admission diagnosis, and comorbidity. In these models, patients with missing data were excluded. Age was modeled with splines using 4 knots.

The Bayesian modelling program Stan was used via the statistical programming language R using the package *brms* [34–36]. Minimally informative priors were used. Models were run for 4000–7500 iterations with 1000–5000 iterations warmup, 4 chains, and 4 cores. Chains, r-hat values and parameter distributions were inspected to assess model fit. Posterior medians with 95% credible intervals were reported. Code and Monte Carlo diagnostics are available in the Supporting information.

Patients with missing surprise question response data were excluded. The impact of missing mortality data were assessed by repeating the mortality analysis with multiple imputation using 25 datasets [37].

Results

Demographic and clinical characteristics

We gathered surprise question responses on 546 patients from 30 attending-trainee physician dyads at 7 hospitals. The median patient age was 75 years (IQR 60–85), 260 (48%) had female sex, median length-of-stay before survey was 8 days (IQR 3–19) and most common admission diagnoses included pneumonia (10%), delirium (9%), and congestive heart failure (9%) (additional details in [Table 1](#)). Three patients (not included in the 546) had incomplete surprise question response data.

Trainee physicians were internal medicine residents in their second or third year of post-graduate training and the median time on service before survey was 18 days (IQR 14–21). Among attending physicians, the median time between survey and graduating medical school was 13 years (IQR 8–23) and the median time on service before survey was 10 days (IQR 9–14).

Surprise question results

Among the 546 patients, attending physicians answered “No, I would not be surprised if this patient died in the next 12 months” (“No”) for 368 patients (67%), while trainees answered “No” for 316 patients (58%). Attending and trainee physicians had the same response of “No” for 272 patients (50%) and the same response of “Yes” for 134 patients (24%) with discordant responses in the remaining 140 patients (26%) ([Table 2](#)).

Twelve-month surprise question responses by attending and trainee physicians showed moderate agreement with a Cohen’s kappa statistic (adjusted for clustering by physician dyad) of 0.54 (95% credible interval 0.41–0.66) for the average physician dyad. The Cohen’s kappa without adjustment for clustering by physician dyad was 0.46 (95% confidence interval 0.38–0.54). In discordant cases, attending physicians were more likely to answer “No” than trainee physicians (relative risk 2.36, 95% credible interval 1.22–4.98). There was moderate variation in kappa across the physician dyads ([Fig 1](#)). Results for the surprise question with respect to hospital discharge are available in the [Supporting information](#).

Associations between agreement and patient covariates

Associations between patient characteristics and physician responses were investigated through a larger adjusted Bayesian multinomial regression model. Two patients were excluded for missing covariate data. Age was modeled with restricted cubic splines using 4 knots and showed a nonlinear relationship between median kappa and age ([Fig 2](#)) with the highest agreement at approximately age 40 and the lowest agreement at approximately age 75. Agreement decreased with presence of a respiratory or cancer comorbidity and increased with presence of infection as admitting diagnosis ([Table 3](#)), although credible intervals were wide throughout the exploratory results.

Comparing surprise question results and mortality

Vital status at 12 months was confirmed for 417 patients (76%). Of these patients, 160 (38%) had died. Across sites, confirmation of mortality status ranged from 47% to 99% and observed confirmed mortality rates ranged from 25% to 53%.

Table 1. Baseline characteristics.

Characteristic	Number (%)
Hospitals	
Total hospitals	7
Patients enrolled per site (median [range])	73 [56–104]
Physician dyads per hospital (median [range])	4 [3–6]
Physicians	
Total dyads	30
Patients per dyad (median [IQR])	18 [16–20]
Attending physician	
Years since medical graduation (median [IQR])	13 [8–23]
Days on current service (median [IQR])	10 [9–14]
Trainee physician	
Years since medical graduation (median [IQR])	2 [1–3]
Days on current service (median [IQR])	18 [14–21]
Patients	
Total patients	546 (100%)
Patient age (years)	
<30	27 (5%)
30–44	32 (6%)
45–59	83 (15%)
60–74	126 (23%)
75–89	232 (42%)
≥ 90	44 (8%)
Missing	2 (0.4%)
Sex	
Female	260 (48%)
Male	284 (52%)
Missing	2 (0.4%)
Residence Type (prior to hospitalization)	
House or Apartment	422 (77%)
Retirement Home or Long-term Care	84 (15%)
No fixed address	19 (3%)
Other	21 (4%)
Function (IADLs/ADLs)	
Independent for all	179 (33%)
Dependent for some	216 (40%)
Unknown	151 (28%)
Admitting diagnoses	
Cardiovascular	78 (14%)
Neurologic	75 (14%)
Infectious	156 (29%)
Acute kidney injury/metabolic abnormality	72 (13%)
Other	165 (30%)
Comorbidities	
Cardiovascular	401 (73%)
Respiratory	88 (16%)
Chronic kidney disease	68 (12%)
Cancer	127 (23%)

(Continued)

Table 1. (Continued)

Characteristic	Number (%)
CPR status	
Full code or not documented	368 (67%)
No CPR	178 (33%)

<https://doi.org/10.1371/journal.pone.0247571.t001>

Among the patients for whom mortality data were available, the probability of death given the surprise question responses was calculated using Bayesian logistic regression accounting for clustering by physician dyad. The probability of death according to surprise question responses was 57% (95% credible interval 49% to 65%) for both attending and trainee physicians responding “No” and 8.3% (95% credible interval 3.7% to 15%) for both attending and trainee physicians responding “Yes.” The probability of death given a response of “No” was 49% (95% credible interval 42% to 56%) for attending physicians and 54% (95% credible interval 46% to 62%) for trainee physicians. Conversely, the probability of death given a response of “Yes” to the surprise question was 14% (95% credible interval 8% to 21%) for attending physicians and 17% (95% credible interval 11% to 23%) for trainee physicians. The analyses found very similar results after multiple imputation of missing data (**Table E1** and **Table E2** in the **Supporting information**).

The adjusted likelihood ratio for death given a response of “No” to the surprise question was 1.84 (95% credible interval 1.55 to 2.22) for trainee physicians and 1.51 (95% credible interval 1.30 to 1.72) for attending physicians. The corresponding negative likelihood ratios were 0.31 (95% credible interval 0.19 to 0.50) and 0.27 (95% confidence interval 0.13 to 0.45). Further details about sensitivity and specificity are available in **Table E3** in the **Supporting information**.

Discussion

This study of 546 general medicine inpatients across 30 attending-trainee physician dyads of physicians showed that attending and trainee physicians had moderate agreement on the surprise question “Would you be surprised if this patient died in the next 12 months?” The adjusted Cohen’s kappa was 0.54, which means that the average dyad agreed on 54% of patients after removing the patients where agreement occurred by chance [30]. The classification of this value of kappa as “moderate” is based on convention [31] and a 54% agreement rate after removing chance agreements is not reassuring in the setting of identifying patients at high risk of mortality and unmet palliative care needs.

The variation across dyads was similar in magnitude to variation by clinical characteristics in exploratory analyses. Agreement by Cohen’s kappa in this study was similar to that seen in

Table 2. 12-month surprise question responses of attending and trainee physicians.

Trainee Physician	Attending Physician			
		“No, I would not be surprised if this patient died in the next 12 months.”	“Yes, I would be surprised if this patient died in the next 12 months.”	
	“No, I would not be surprised if this patient died in the next 12 months.”	272 (49%)	44 (8%)	316 (57%)
	“Yes, I would be surprised if this patient died in the next 12 months.”	96 (18%)	134 (25%)	230 (43%)
		368 (67%)	178 (33%)	546 (100%)

<https://doi.org/10.1371/journal.pone.0247571.t002>

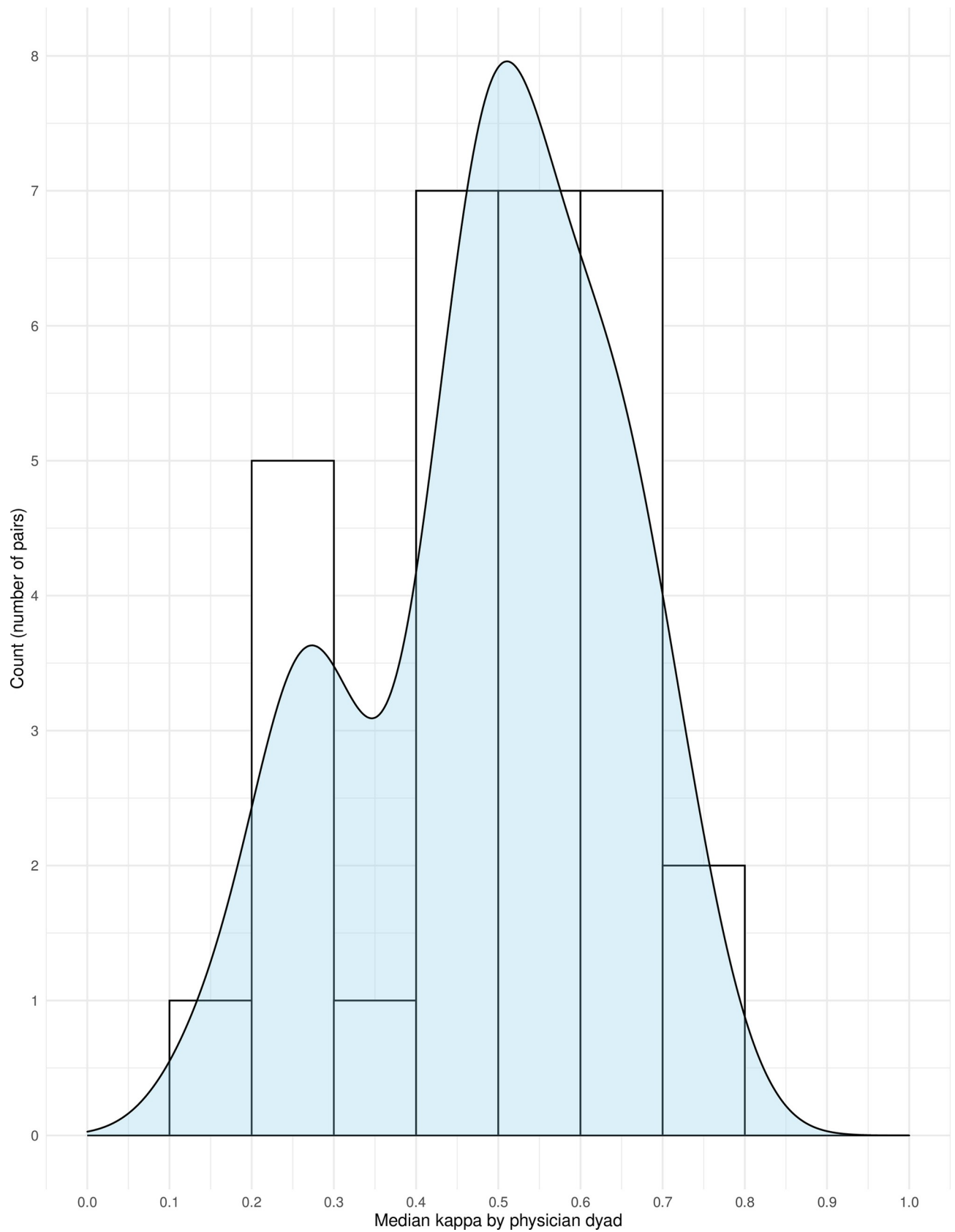


Fig 1. Histogram and density plot of median kappa values by physician dyad. This figure shows the distribution of kappa values according to physician dyad by histogram (bars) with a density plot overlay (light blue). Most clinician dyads had a kappa value above 0.4, but some outliers had kappa values between 0.1 and 0.3.

<https://doi.org/10.1371/journal.pone.0247571.g001>

another study [15] comparing predictions of nurses and physicians, suggesting that moderate agreement on prognosis may be common to other dyads of clinicians beyond trainee-attending dyads. Taken together, these findings imply that discussion of patient prognosis between multidisciplinary clinical team members is essential to ensure appropriate engagement of palliative care services and to avoid mixed messages to patients and family members.

The surprise question may be useful to “rule out” a high risk of mortality, but it is not sufficient as a standalone screening measure for identifying patients who have high risk of mortality. A response of “Yes” on the surprise question was associated with low 12-month mortality in this and other settings, but the sensitivity and specificity remain inconsistent and unsuited to a screening tool [2,38]. The lowest 12-month mortality rate was seen in patients where both attending and trainee physician responded “Yes” to the surprise question, similar to other studies where combining predictions of multiple clinical team members yielded the best predictions [15]. In contrast to previous studies investigating physician-estimated prognoses which found correlation between accuracy and level of training [11,39–42], attending physician predictions were not more accurate than trainee physician predictions with respect to mortality in our study. This could mean that both attending and trainee physicians would benefit from educational interventions focused on prognosis. Alternatively, accurate prognostication may require clinicians to combine clinical insights with novel tools not yet in clinical use, such as automated screening tools derived from electronic medical records [43–46].

This study has strengths including a pragmatic approach, inclusive enrollment criteria, paired design, multicenter data, and statistical methods that account for clustering. The simplicity of the surprise question allowed us to include all patients within our inclusion criteria with no planned or unplanned systematic exclusions on the day of survey. Our paired design minimized patient-level confounding and comparing to the most senior trainee as comparator minimized error due to medical inexperience. The choice of Bayesian modeling permitted a more sophisticated analysis of the Cohen’s kappa coefficient which has not previously been used in analyses of the surprise question [2,38].

The main limitation of this study is the lack of data on unmet palliative care needs such as uncontrolled pain or nausea, psychological distress about the dying process, or ignorance about the available options for end-of-life care. This limitation is also present in other research on the surprise question [22]. It is unknown whether the attending or trainee physician is more likely to be correct with respect to palliative care needs when there is discordance. A related limitation is a lack of corresponding qualitative data including information about the reasoning behind the responses of each participant. It is also unknown whether the two physicians were aware when they did not agree, or if this was discussed for any of the patients in the study.

A more fundamental limitation is that the surprise question is a subjective instrument that integrates a healthcare practitioner’s expertise, knowledge, and personal biases [47]. This study includes uncertainty both from the intrinsic uncertainty in estimating a patients’ prognosis and from the uncertainty in how clinicians interpret this uncertainty in responding to the surprise question. The validity of the surprise question with respect to its intended purpose of identifying patients with palliative care needs remains unclear. Even if the surprise question reliably identifies those patients, identification alone is necessary but not sufficient for meeting those needs [48,49]. Future research may need to focus on identification of unmet palliative

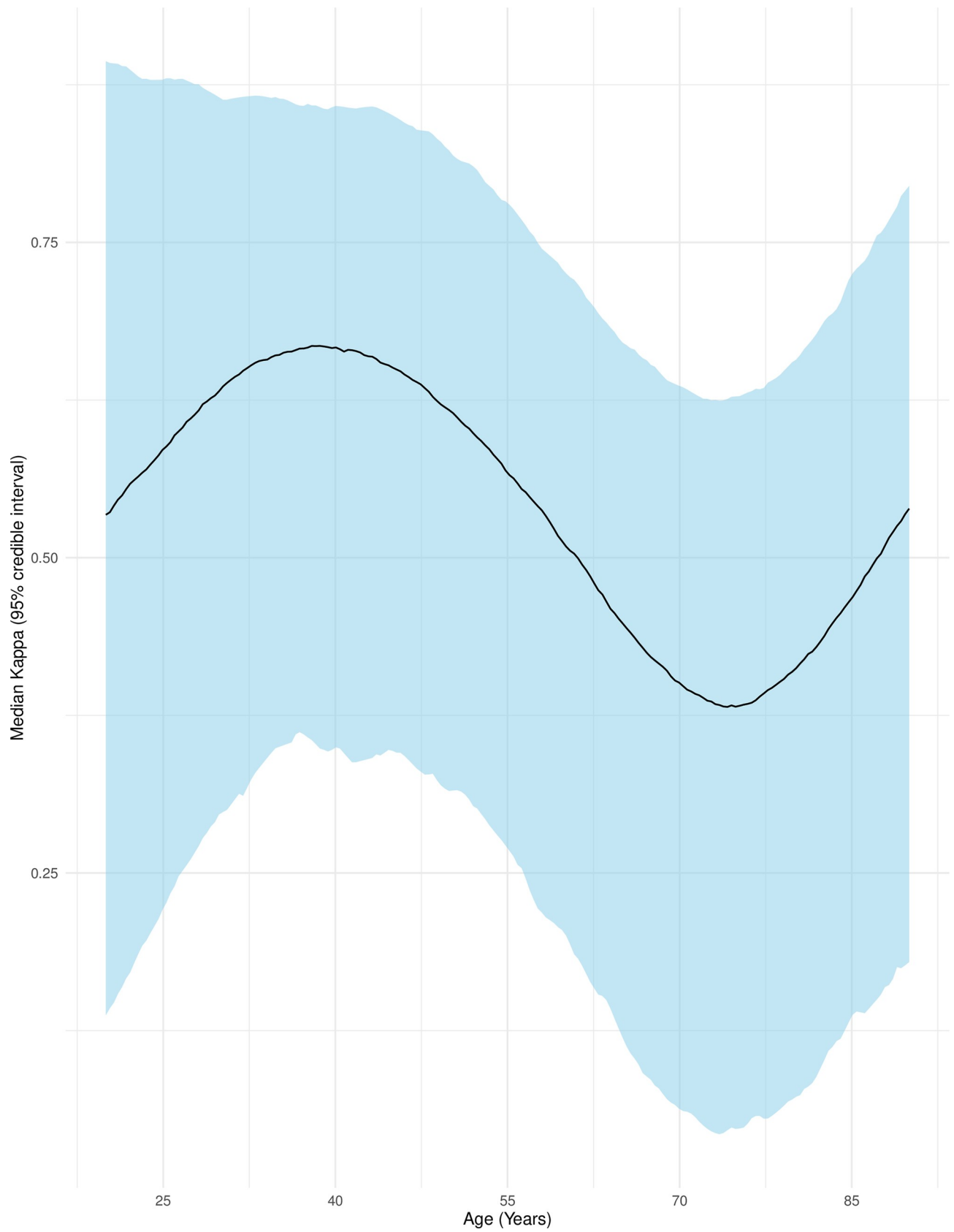


Fig 2. Agreement between attending and trainee physician surprise question responses by patient age. This figure shows the median Cohen's kappa (adjusted for clustering by physician dyad) as a black line with 95% credible intervals as surrounding blue ribbon. Age was modeled with restricted cubic splines using 4 knots. Other variables were set to: Female, independent, full code, infection as admitting diagnosis, and presence of a cardiovascular comorbidity only.

<https://doi.org/10.1371/journal.pone.0247571.g002>

care needs as opposed to predicting mortality, as these needs may be a better marker of potential benefit from a palliative approach and provide tangible targets for the clinical team.

Conclusion

Using the surprise question to measure prognosis, the average general internal medicine attending and trainee physician agreed on the prognosis of only 54% of their patients after correcting for chance agreement. These data remind clinicians of the subjectivity in formulating patient prognoses, and the importance of routine discussion of patient prognosis between team members in order to ensure a coherent clinical plan including clear communication with patients and appropriate engagement of palliative care services.

Table 3. Agreement between attending and trainee physician by patient subgroup, adjusted for multiple clinical covariates.

Patient characteristic	Median Kappa (95% credible interval)
Age (years)	
30	0.69 (0.34 to 0.89)
45	0.72 (0.41 to 0.90)
60	0.59 (0.27 to 0.82)
75	0.47 (0.14 to 0.74)
90	0.64 (0.31 to 0.87)
Sex	
Female	0.59 (0.26 to 0.81)
Male	0.61 (0.35 to 0.84)
Function (IADLs/ADLs)	
Independent for all	0.59 (0.28 to 0.83)
Dependent for some	0.55 (0.19 to 0.82)
Unknown	0.56 (0.16 to 0.80)
CPR status	
Full code and not documented	0.57 (0.28 to 0.83)
Not for CPR	0.46 (0.02 to 0.78)
Admitting diagnosis	
Cardiovascular	0.59 (0.29 to 0.81)
Neurologic	0.50 (0.21 to 0.73)
Infectious	0.71 (0.42 to 0.87)
AKI/metabolic	0.41 (-0.07 to 0.75)
Other	0.31 (-0.12 to 0.67)
Comorbidities	
Cardiovascular	0.46 (0.06 to 0.80)
Respiratory	0.17 (-0.20 to 0.64)
Chronic kidney disease	0.42 (0.00 to 0.82)
Cancer	0.29 (-0.04 to 0.75)

Unless otherwise noted, the baseline characteristics used for calculating median kappas and 95% credible intervals were: Age 60 years, female, independent, full code, cardiovascular admitting diagnosis and cardiovascular comorbidity.

<https://doi.org/10.1371/journal.pone.0247571.t003>

Supporting information

S1 File.

(DOCX)

S2 File.

(DOCX)

S3 File.

(CSV)

Acknowledgments

An appreciative thanks to the following people: Carol Mantle and Marilyn Swinton for assistance with data collection and project coordination at Hamilton Health Sciences. Katherine Allan for assistance with data collection and project coordination at St Michael's Hospital. Ellen Koo for assistance with data collection and project coordination at Toronto General and Toronto Western Hospitals. Nicole Marinoff for assistance with project coordination at Sunnybrook Health Sciences Center. Julia Kruizinga for assistance with data collection at Kingston Health Science Centre.

Author Contributions

Conceptualization: Christopher J. Yarnell, Laura M. Jewell, Ruxandra Pinto, Luke A. Devine, James Downar, Roy Ilan, Shail Rawal, Natalie Wong, John J. You, Rob A. Fowler.

Data curation: Christopher J. Yarnell, Laura M. Jewell, Alex Astell, Luke A. Devine, James Downar, Roy Ilan, Shail Rawal, Natalie Wong, John J. You, Rob A. Fowler.

Formal analysis: Christopher J. Yarnell, Laura M. Jewell, Ruxandra Pinto, Michael E. Detsky, Natalie Wong, John J. You, Rob A. Fowler.

Funding acquisition: Christopher J. Yarnell, Laura M. Jewell, Rob A. Fowler.

Investigation: Christopher J. Yarnell, Laura M. Jewell, Alex Astell, James Downar, Roy Ilan, Shail Rawal, Natalie Wong, John J. You, Rob A. Fowler.

Methodology: Christopher J. Yarnell, Laura M. Jewell, Alex Astell, Ruxandra Pinto, James Downar, Roy Ilan, Shail Rawal, Natalie Wong, John J. You, Rob A. Fowler.

Project administration: Christopher J. Yarnell, Laura M. Jewell, Alex Astell, Luke A. Devine, James Downar, Roy Ilan, Shail Rawal, Natalie Wong, John J. You, Rob A. Fowler.

Resources: Christopher J. Yarnell, Laura M. Jewell, Alex Astell, Luke A. Devine, James Downar, Roy Ilan, Shail Rawal, Natalie Wong, John J. You, Rob A. Fowler.

Software: Christopher J. Yarnell.

Supervision: Rob A. Fowler.

Visualization: Christopher J. Yarnell, Rob A. Fowler.

Writing – original draft: Christopher J. Yarnell.

Writing – review & editing: Christopher J. Yarnell, Laura M. Jewell, Alex Astell, Ruxandra Pinto, Luke A. Devine, Michael E. Detsky, James Downar, Roy Ilan, Shail Rawal, Natalie Wong, John J. You, Rob A. Fowler.

References

1. Yourman LC, Lee SJ, Schonberg MA, Widera EW, Smith AK. Prognostic indices for older adults: a systematic review. *JAMA* [Internet]. 2012 Jan 11 [cited 2019 Jun 19]; 307(2):182–92. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22235089>. <https://doi.org/10.1001/jama.2011.1966> PMID: 22235089
2. Downar J, Goldman R, Pinto R, Englesakis M, Adhikari NKJ. The “surprise question” for predicting death in seriously ill patients: a systematic review and meta-analysis. *CMAJ* [Internet]. 2017 Apr 3 [cited 2019 Jun 19]; 189(13):E484–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/28385893>. <https://doi.org/10.1503/cmaj.160775> PMID: 28385893
3. Vincent J-L, Moreno R. Clinical review: Scoring systems in the critically ill. *Crit Care* [Internet]. 2010 [cited 2019 Jun 19]; 14(2):207. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20392287>. <https://doi.org/10.1186/cc8204> PMID: 20392287
4. Vincent J-L, Opal SM, Marshall JC. Ten reasons why we should NOT use severity scores as entry criteria for clinical trials or in our treatment decisions*. *Crit Care Med* [Internet]. 2010 Jan [cited 2019 Jun 19]; 38(1):283–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/19730252>. <https://doi.org/10.1097/CCM.0b013e3181b785a2> PMID: 19730252
5. van Walraven C, McAlister FA, Bakal JA, Hawken S, Donzé J. External validation of the Hospital-patient One-year Mortality Risk (HOMR) model for predicting death within 1 year after hospital admission. *CMAJ* [Internet]. 2015 Jul 14 [cited 2019 Jun 20]; 187(10):725–33. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/26054605>. <https://doi.org/10.1503/cmaj.150209> PMID: 26054605
6. Murray SA, Kendall M, Boyd K, Sheikh A. Illness trajectories and palliative care. *BMJ* [Internet]. 2005 Apr 30 [cited 2019 Jun 19]; 330(7498):1007–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15860828>. <https://doi.org/10.1136/bmj.330.7498.1007> PMID: 15860828
7. Koffman J, Yorganci E, Yi D, Gao W, Murtagh F, Pickles A, et al. Managing uncertain recovery for patients nearing the end of life in hospital: a mixed-methods feasibility cluster randomised controlled trial of the AMBER care bundle. *Trials* [Internet]. 2019; 20(1):506. Available from: <https://doi.org/10.1186/s13063-019-3612-0> PMID: 31419994
8. Stone P, Gwilliam B, Keeley V, Todd C, Gittins M, Kelly L, et al. Patients reports or clinicians assessments: which are better for prognosticating? *BMJ Support Palliat Care* [Internet]. 2012; 2(3):219–23. Available from: <https://spcare.bmj.com/content/2/3/219>. <https://doi.org/10.1136/bmjspcare-2012-000216> PMID: 24654194
9. Downar J, Wegier P, Tanuseputro P. Early Identification of People Who Would Benefit From a Palliative Approach—Moving From Surprise to Routine. *JAMA Netw open*. 2019 Sep 4; 2(9):e1911146. <https://doi.org/10.1001/jamanetworkopen.2019.11146> PMID: 31517959
10. Goettler CE, Waibel BH, Goodwin J, Watkins F, Toschlog EA, Sagraves SG, et al. Trauma intensive care unit survival: how good is an educated guess? *J Trauma* [Internet]. 2010 Jun [cited 2019 Jun 19]; 68(6):1279–87; discussion 1287–8. Available from: <https://insights.ovid.com/crossref?an=00005373-201006000-00002>. <https://doi.org/10.1097/TA.0b013e3181de3b99> PMID: 20539170
11. Barrera R, Nygard S, Sogoloff H, Groeger J, Wilson R. Accuracy of predictions of survival at admission to the intensive care unit. *J Crit Care* [Internet]. 2001 Mar [cited 2019 Jun 19]; 16(1):32–5. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/11230722>. <https://doi.org/10.1053/jcrc.2001.21794> PMID: 11230722
12. Hui D, Kilgore K, Nguyen L, Hall S, Fajardo J, Cox-Miller TP, et al. The accuracy of probabilistic versus temporal clinician prediction of survival for patients with advanced cancer: a preliminary report. *Oncologist* [Internet]. 2011 [cited 2019 Jun 19]; 16(11):1642–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21976316>. <https://doi.org/10.1634/theoncologist.2011-0173> PMID: 21976316
13. Glare PA, Sinclair CT. Palliative Medicine Review: Prognostication. *J Palliat Med* [Internet]. 2008 Jan [cited 2019 Jun 19]; 11(1):84–103. Available from: <http://www.liebertpub.com/doi/10.1089/jpm.2008.9992>. <https://doi.org/10.1089/jpm.2008.9992> PMID: 18370898
14. Poses RM, Bekes C, Winkler RL, Scott WE, Copare FJ. Are two (inexperienced) heads better than one (experienced) head? Averaging house officers’ prognostic judgments for critically ill patients. *Arch Intern Med* [Internet]. 1990 Sep [cited 2019 Jun 19]; 150(9):1874–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2393319>. PMID: 2393319
15. Detsky ME, Harhay MO, Bayard DF, Delman AM, Buehler AE, Kent SA, et al. Discriminative accuracy of physician and nurse predictions for survival and functional outcomes 6 months after an ICU admission. *JAMA—J Am Med Assoc*. 2017 Jun 6; 317(21):2187–95. <https://doi.org/10.1001/jama.2017.4078> PMID: 28528347
16. Heyland DK, Barwich D, Pichora D, Dodek P, Lamontagne F, You JJ, et al. Failure to Engage Hospitalized Elderly Patients and Their Families in Advance Care Planning. *JAMA Intern Med* [Internet]. 2013 May 13 [cited 2019 Jun 19]; 173(9):778. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23545563>. <https://doi.org/10.1001/jamainternmed.2013.180> PMID: 23545563

17. A controlled trial to improve care for seriously ill hospitalized patients. The study to understand prognoses and preferences for outcomes and risks of treatments (SUPPORT). The SUPPORT Principal Investigators. *JAMA* [Internet]. [cited 2019 Jun 19]; 274(20):1591–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/7474243>.
18. Moss AH, Ganjoo J, Sharma S, Gansor J, Senft S, Weaner B, et al. Utility of the “Surprise” Question to Identify Dialysis Patients with High Mortality. *Clin J Am Soc Nephrol* [Internet]. 2008 Sep [cited 2019 Jun 19]; 3(5):1379–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18596118>. <https://doi.org/10.2215/CJN.00940208> PMID: 18596118
19. Da Silva Gane M, Braun A, Stott D, Wellsted D, Farrington K. How Robust Is the “Surprise Question” in Predicting Short-Term Mortality Risk in Haemodialysis Patients. *Nephron Clin Pract* [Internet]. 2013 [cited 2019 Jun 19]; 123(3–4):185–93. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23921223>. <https://doi.org/10.1159/000353735> PMID: 23921223
20. Thomas K, Free A, GSF Central Team. The Gold Standards Framework Prognostic Indicator Guidance [Internet]. 2011. Available from: <http://www.goldstandardsframework.org.uk/cd-content/uploads/files/GeneralFiles/PrognosticIndicatorGuidanceOctober2011.pdf>.
21. Pang W-F, Kwan BC-H, Chow K-M, Leung C-B, Li PK-T, Szeto C-C. Predicting 12-month mortality for peritoneal dialysis patients using the “surprise” question. *Perit Dial Int* [Internet]. 2013 [cited 2019 Jun 19]; 33(1):60–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22855890>. <https://doi.org/10.3747/pdi.2011.00204> PMID: 22855890
22. Romo RD, Lynn J. The utility and value of the “surprise question” for patients with serious illness. Vol. 189, *CMAJ. Canadian Medical Association*; 2017. p. E1072–3. <https://doi.org/10.1503/cmaj.733231> PMID: 28827439
23. Lee TH, Marcantonio ER, Mangione CM, Thomas EJ, Polanczyk CA, Cook EF, et al. Derivation and prospective validation of a simple index for prediction of cardiac risk of major noncardiac surgery. *Circulation* [Internet]. 1999 Sep 7 [cited 2019 Jun 19]; 100(10):1043–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10477528>. <https://doi.org/10.1161/01.cir.100.10.1043> PMID: 10477528
24. Dy S, Lynn J. Getting services right for those sick enough to die. *BMJ* [Internet]. 2007 Mar 10 [cited 2019 Jun 19]; 334(7592):511–3. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17347238>. <https://doi.org/10.1136/bmj.39127.653704.80> PMID: 17347238
25. Hill AD, Stukel TA, Fu L, Scales DC, Laupacis A, Rubenfeld GD, et al. Trends in site of death and health care utilization at the end of life: a population-based cohort study. *C Open* [Internet]. 2019 Apr 26 [cited 2019 Jun 19]; 7(2):E306–15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/31028054>. <https://doi.org/10.9778/cmajo.20180097> PMID: 31028054
26. Bekelman JE, Halpern SD, Blankart CR, Bynum JP, Cohen J, Fowler R, et al. Comparison of Site of Death, Health Care Utilization, and Hospital Expenditures for Patients Dying With Cancer in 7 Developed Countries. *JAMA* [Internet]. 2016 Jan 19 [cited 2019 Jun 19]; 315(3):272. Available from: <http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.18603>. PMID: 26784775
27. Quinn KL, Stall NM, Yao Z, Stukel TA, Cram P, Detsky AS, et al. The risk of death within 5 years of first hospital admission in older adults. *CMAJ*. 2019 Dec 16; 191(50):E1369–77. <https://doi.org/10.1503/cmaj.190770> PMID: 31844021
28. Royal College of Physicians and Surgeons Canada. Objectives of Training in the Specialty of Internal Medicine [Internet]. 2015. Available from: https://www.royalcollege.ca/rcsite/documents/ibd/internal_medicine_otr_e.pdf.
29. Koffman J, Morgan M, Edmonds P, Speck P, Higginson IJ. Vulnerability in palliative care research: findings from a qualitative study of black Caribbean and white British patients with advanced cancer. *J Med Ethics*. 2009 Jul; 35(7):440–4. <https://doi.org/10.1136/jme.2008.027839> PMID: 19567695
30. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas* [Internet]. 1960 Apr 2 [cited 2020 May 30]; 20(1):37–46. Available from: <http://journals.sagepub.com/doi/10.1177/001316446002000104>.
31. Landis JR, Koch GG. The Measurement of Observer Agreement for Categorical Data. *Biometrics* [Internet]. 1977 Mar [cited 2020 May 30]; 33(1):159. Available from: <https://www.jstor.org/stable/2529310>. PMID: 843571
32. Hsiao CK, Chen PC, Kao WH. Bayesian random effects for interrater and test-retest reliability with nested clinical observations. *J Clin Epidemiol*. 2011 Jul; 64(7):808–14. <https://doi.org/10.1016/j.jclinepi.2010.10.015> PMID: 21292442
33. Fitzmaurice Garrett M., Laird Nan M., Ware James H. Applied longitudinal analysis/Garrett M. Fitzmaurice, Nan M. Laird, James H. Ware [Internet]. Wiley-Interscience Hoboken, N.J.; 2004. xvii, 506 p.: Available from: <http://www.loc.gov/catdir/toc/wiley041/2004040891.html>.
34. Bürkner PC. brms: An R package for Bayesian multilevel models using Stan. *J Stat Softw*. 2017 Aug 29; 80(1):1–28.

35. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: <https://www.r-project.org>.
36. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. *J Stat Softw* [Internet]. 2017 Jan 11 [cited 2020 Jul 2]; 76(1):1–32. Available from: <https://www.jstatsoft.org/index.php/jss/article/view/v076i01/v76i01.pdf>.
37. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci*. 2007 Sep; 8(3):206–13. <https://doi.org/10.1007/s11121-007-0070-9> PMID: 17549635
38. White N, Kupeli N, Vickerstaff V, Stone P. How accurate is the “Surprise Question” at identifying patients at the end of life? A systematic review and meta-analysis. *BMC Med*. 2017 Aug 2; 15(1). <https://doi.org/10.1186/s12916-017-0907-4> PMID: 28764757
39. Poses RM, Bekes C, Copare FJ, Scott WE. The answer to “What are my chances, doctor?” depends on whom is asked: prognostic disagreement and inaccuracy for critically ill patients. *Crit Care Med* [Internet]. 1989 Aug [cited 2019 Jun 19]; 17(8):827–33. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/2752779>.
40. Poses RM, McClish DK, Bekes C, Scott WE, Morley JN. Ego bias, reverse ego bias, and physicians’ prognostic. *Crit Care Med* [Internet]. 1991 Dec [cited 2019 Jun 19]; 19(12):1533–9. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/1959374>. <https://doi.org/10.1097/00003246-199112000-00016> PMID: 1959374
41. Christakis NA, Lamont EB. Extent and determinants of error in doctors’ prognoses in terminally ill patients: prospective cohort study. *BMJ* [Internet]. 2000 Feb 19 [cited 2019 Jun 19]; 320(7233):469–72. Available from: <http://www.bmj.com/cgi/doi/10.1136/bmj.320.7233.469>. <https://doi.org/10.1136/bmj.320.7233.469> PMID: 10678857
42. Frick S, Uehlinger DE, Zuercher Zenklusen RM. Medical futility: Predicting outcome of intensive care unit patients by nurses and doctors—A prospective comparative study*. *Crit Care Med* [Internet]. 2003 Feb [cited 2019 Jun 19]; 31(2):456–61. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12576951>. <https://doi.org/10.1097/01.CCM.0000049945.69373.7C> PMID: 12576951
43. Wegier P, Koo E, Ansari S, Kobewka D, O’connor E, Wu P, et al. MHOMR: a feasibility study of an automated system for identifying inpatients having an elevated risk of 1-year mortality. *BMJ Qual Saf*. 2019 Dec 1; 28(12):971–9. <https://doi.org/10.1136/bmjqs-2018-009285> PMID: 31253736
44. Courtright KR, Chivers C, Becker M, Regli SH, Pepper LC, Draugelis ME, et al. Electronic Health Record Mortality Prediction Model for Targeted Palliative Care Among Hospitalized Medical Patients: a Pilot Quasi-experimental Study. *J Gen Intern Med*. 2019 Sep 15; 34(9):1841–7. <https://doi.org/10.1007/s11606-019-05169-2> PMID: 31313110
45. Hsu AT, Manuel DG, Taljaard M, Chalifoux M, Bennett C, Costa AP, et al. Algorithm for predicting death among older adults in the home care setting: Study protocol for the Risk Evaluation for Support: Predictions for Elder-life in the Community Tool (RESPECT). *BMJ Open*. 2016 Dec 1; 6(12).
46. Wegier P, Steinberg L, Myers J, Koo E, Saunders S, Kurahashi A, et al. MHOMR: AN AUTOMATED MORTALITY PREDICTION TOOL TO IDENTIFY PATIENTS WITH UNMET PALLIATIVE CARE-RELATED NEEDS. In: 41st Annual Meeting of the Society for Medical Decision Making. Portland, Oregon; 2020. p. PS 1–22.
47. Elliott M, Nicholson C. A qualitative study exploring use of the surprise question in the care of older people: perceptions of general practitioners and challenges for practice. *BMJ Support Palliat Care* [Internet]. 2017 Mar [cited 2019 Jun 19]; 7(1):32–8. Available from: <http://spcare.bmj.com/lookup/doi/10.1136/bmjspcare-2014-000679>. <https://doi.org/10.1136/bmjspcare-2014-000679> PMID: 25168076
48. You JJ, Downar J, Fowler RA, Lamontagne F, Ma IW, Jayaraman D, et al. Barriers to goals of care discussions with seriously ill hospitalized patients and their families: a multicenter survey of clinicians. [Erratum appears in *JAMA Intern Med*. 2015 Apr; 175(4):659; <https://doi.org/10.1001/jamainternmed.2014.7732> PMID: 25642797]. *JAMA Intern Med* [Internet]. 2015 [cited 2020 Feb 3]; 175:549–56. Available from: <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&CSC=Y&NEWS=N&PAGE=fulltext&D=pem&AN=25642797http://oxfordfx.hosted.exlibrisgroup.com/oxford?sid=OVID:medline&id=pmid:25642797&id=doi:10.1001%2Fjamainternmed.2014.7732&issn=2168-6106&isbn=&volume=175&issue=4&spage=>.
49. Hudson KE, Wolf SP, Samsa GP, Kamal AH, Abernethy AP, Leblanc TW. The surprise question and identification of palliative care needs among hospitalized patients with advanced hematologic or solid malignancies. *J Palliat Med* [Internet]. 2018 Jun 1 [cited 2020 Mar 11]; 21(6):789–95. Available from: <http://www.liebertpub.com/doi/10.1089/jpm.2017.0509>. <https://doi.org/10.1089/jpm.2017.0509> PMID: 29420142