BMJ Open

How accurate is clinical prognostication by oncologists during routine practice in a general hospital and can it be improved by a specific prognosis training programme: a prospective interventional study

Irma Kupf ,[1] Gabriele Thanner,[2] Michael Gerken,[3] Alexander Crispin,[4] Jan Braess[5]

To cite: Kupf I, Thanner G, Gerken M, *et al*. How accurate is clinical prognostication by oncologists during routine practice in a general hospital and can it be improved by a specific prognosis training programme: a prospective interventional study. *BMJ Open* 2024;**14**:e081661. doi:10.1136/bmjopen-2023-081661

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2023-081661).

Received 02 November 2023
Accepted 20 May 2024

For numbered affiliations see end of article.

Correspondence to
Professor Jan Braess;
Jan.Braess@barmherzige-regensburg.de

ABSTRACT

Objectives Oncologists need competence in clinical prognostication to deliver appropriate care to patients with cancer. Most studies on prognostication have been restricted to patients in palliative care settings. This paper investigates (1) the prognostic accuracy of physicians regarding a broad cohort of patients with cancer with a median life expectancy of >2 years and (2) whether a prognosis training can improve prognostication.

Design Prospective single-centre study comprising 3 phases, each lasting 1 month.

Setting Large teaching hospital, department of oncology and haematology, Germany.

Participants 18 physicians with a professional experience from entry level to 34 years. 736 patients with oncological and malignant haematological diseases.

Interventions Baseline prognostication abilities were recorded during an 'untrained' phase 1. As an intervention, a specific prognosis-training programme was implemented prior to phases 2 and 3. In phase 3, physicians had to provide additional estimates with the inclusion of electronic prognostic tools.

Outcome measures Prognostic estimates (PE) were collected using 'standard' surprise question (SQ), 'probabilistic' SQ (both for short-term prognostication up to 6 months) and clinician prediction of survival (CPS) (for long-term prognostication). Estimated prognoses were compared with observed survival. Phase 1 was compared with phases 2 and 3.

Results We included 2427 PE for SQ, 1506 for CPS and 800 for probabilistic SQ. Median OS was 2.5 years. SQ accuracy improved significantly (p<0.001) from 72.6% in phase 1 to 84.3% in phase 3. Probabilistic SQ in phase 3 showed 83.1% accuracy. CPS accuracy was 25.9% and could not be significantly improved. (Electronic) prognostic tools—used alone—performed significantly worse (p<0.0005) than physicians and—used by the clinicians—did not improve their performance.

Conclusion A specific prognosis-training programme could improve short-term and intermediate-term prognostication. Improvement of long-term prognostication was not possible. Inexperienced residents as well as experienced oncologists benefited from training.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This is the first study that presents a structured training programme for oncologists that has a positive impact on clinical prognostication.
⇒ This prospective study included a large cohort of patients with different oncological and malignant haematological diseases at all disease stages.
⇒ Prospective follow-up data of almost 4 years were collected, which allowed for the assessment of longer prognosis estimates.
⇒ The effect of training was not demonstrated by a randomised comparison but by comparing baseline (untrained) prognostication abilities with the performance after training in the same group of physicians.
⇒ The study is a single-centre study—even though performed at a large teaching hospital.

INTRODUCTION

In order to deliver appropriate care to patients with cancer, an oncologist needs competence not only as a diagnostician or as a therapist but also as a prognosticator. The relevance of prognostication may not always be obvious, but it underlies virtually every aspect of patient care. For example, the same clinical symptom (severe respiratory insufficiency) may lead to very different courses of action (eg, transfer to the ICU vs best supportive care measures) depending on the estimate of the patient's prognosis by the treating physician. The oncologist may have made these estimates consciously or unconsciously, but they nevertheless constitute the

background in front of which everyday patient care takes place.[1]

Prognostication is an uncertain business and the ability of oncologists to estimate the future course of their patients is limited. Most explicit analyses of physicians' abilities to prognosticate have been made in the palliative care setting on patients with very short life expectancies.[2–29] In contrast, this paper evaluates the prognostic abilities of residents and experienced oncologists in a department of oncology and haematology in a major acute care teaching hospital, where patients had an average life expectancy of >2 years. This paper also investigates how far a training programme can improve prognostic performance.

There are three major approaches to indicate the prognosis of a patient: (1) the temporal approach (also called 'clinician prediction of survival' (CPS)), (2) the surprise question (SQ) and (3) the probabilistic approach. Of these, the temporal approach is the most commonly used method within the literature.[29–32]

When using the temporal approach, the prognosticator is required to state a specific period in absolute numbers (days, months, years). Many studies consider the prognosis to be 'accurate' if the CPS is within the range 67%–133% in relation to the actual survival of the observed patient (which is set as 100%).[6 7 11 13 15 16 18 21] The SQ, which was originally developed as a screening tool for the initiation of palliative care, requires physicians to answer whether they would be surprised if their patient had died within a certain time. There are only two options (yes/no) and the estimation period is defined (eg, 6 months). The SQ is accurate if the answer is 'yes' and the patient has survived or if the option is 'no' and the patient has died. The third option is the probabilistic approach. Here, the respondent must indicate the probability that a patient has or has not died at a given time. (Like in the SQ) the period of interest is defined (eg, 6 months), but the prognostic estimate is provided in percentage increments (usually 10%) and thus has a quantitative component.

Prognostication per se (= estimating a prognosis) is not the same as conveying a prognosis to a patient (eg, 'breaking bad news'). The latter ability (which is not the topic of this paper) is also of high relevance and is generally regarded to require considerable experience and clinical expertise (which is considered teachable). In contrast, the preceding process of estimating the prognosis—that is, prognostication as such—is described as 'Medicine's Lost Art' and remains a black box in two ways: (1) The prognostic performance of clinicians—outside the inpatient palliative care setting—is often unknown. (2) Training of prognostication is often 'implicit'—if it happens at all.[33–35] This is in contrast to other fields outside clinical medicine where systematic training approaches to increase prognostic performance and accuracy do exist.[36 37]

For patients in routine oncological practice (and not in the palliative care setting)—no data on improving prognostication via a training programme exist to our knowledge. In this paper, we, therefore, evaluate not only the prognostic performance of residents and experienced oncologists in the acute care setting but also how far a structured training programme might have a positive impact on prognostication.

METHODS

Time period and general conditions

The prospective study was conducted at the clinic for oncology and haematology at the Hospital Barmherzige Brüder in Regensburg, Germany. It consisted of 3 different phases with a duration of 1 month each, taking place in the 3 months of April, June and September of 2019. Data collection was performed as part of routine care on the oncology wards and in the oncology outpatient clinic. All physicians agreed to participate in the study and gave their informed consent. Since the study did not involve any intervention affecting the patient and the standard of care was not affected no informed consent was required for patients. The current study was carried out in accordance with the Declaration of Helsinki.

Participants

The participants of the study consisted of the whole medical team (= physicians) of the department of oncology and haematology (n=21), who (individually) generated the prognostic estimates on the patients (described below). This 'whole group' was subdivided into 'residents' (physicians in training, less than 1 year of training in clinical oncology = 'inexperienced') and into 'oncologists' (board-certified oncologists with a minimum experience in clinical oncology of 13+years = 'experienced').

Procedure

General procedure

Patients who were discharged from the department of oncology and haematology during the 3 phases were included. All patients were required to have a malignant disease but could be in any stage of their disease. After each patient's discharge, the resident physician who last cared for the patient and the supervising senior physician and/or chief physician independently evaluated the patient's prognosis based on knowing the patient personally and based on the information from the discharge letter. To determine both a qualitative value and a quantitative value for the prognostic accuracy, questionnaires were used to elicit the physicians' responses to the following two questions:

1. *'Would you be surprised, if your patient had died in 6 months from today?'* ('surprise question' (SQ)).
2. *'If you were faced with a large group of patients with a very similar constellation of age, disease, comorbidities, prior therapies as in the patient whose discharge letter you have just completed, what median survival would you expect in this group?'* ('clinician prediction of survival' (CPS)).

Sample size calculation/considerations

We determined that each of the 3 phases was to consist of 4 consecutive weeks. Our calculation was that within this period of 4 weeks approximately 300 patients would leave the oncology department and could, therefore, be evaluated for prognosis by their treating physicians. Since the resident as well as one or two senior oncology consultants gave their prognosis estimates (PE) independently on the same patient a total number of approximately 600–900 prognoses per phase were expected (and de facto achieved—phase 1=925, phase 2=712, phase 3=790). We assumed that these rather high numbers (as compared with those within the literature of clinical prognostication) were needed to give an accurate descriptive picture of the status quo on 'untrained' clinical prognostication (in phase 1) and of prognostic performance after training (phases 2 and 3).

Data collection

Every physician's estimate and every hospitalisation of a patient were considered. When there were multiple estimates for a patient (eg, by resident, senior physician and chief physician), all three estimates were recorded as distinct data points for this respective patient for this hospital stay. Patients, who received a PE on more than one occasion, were included in the analysis for each of their PE. For analysis all physicians together formed the 'whole group'. In addition, according to the clinical experience, we defined the subgroups 'residents' and 'oncologists' (board-certified specialists in haematology and oncology).

To avoid the possibility of non-representative values caused by outliers, only physicians who provided ≥10 PE per phase were included in the final analysis. Physicians, who did not work in the oncology wards in every phase (eg, due to rotation in residency training), were included only in the involved phase(s) if they had completed ≥10 PE. Real survival data (date of death or last recorded life status) were provided by the Tumor Center Regensburg in March 2023, with a follow-up of approximately 4 years. PE for patients, who did not have an oncological disease or who were not registered at the Tumor Center Regensburg, were not included in the analysis (because follow-up could not be ensured).

Phase 1 versus phases 2 and 3/interventions

A flow chart of the 3 phases with all participating physicians and patients is provided in online supplemental file 1. In phase 1, the PE was performed with no prior training and was thus based mainly on prior clinical experience and 'gut feeling' ('thinking fast'). In contrast, during phase 2 'thinking slow' (according to Kahneman[38]) was encouraged—that is to consciously take time for the PE. To enable that, a prognosis training was provided for the physicians prior to the start of the second phase and repeated before phase 3 (see online supplemental file 2):

1. The prognosis training consisted of a brochure designed to support the physicians with the aid of epidemiological data on the 21 most common oncological diseases of the clinic. For each disease, general information was provided on the median survival time across all groups as well as information for different patient groups (eg, according to patients' age, stage or mutation status). Corresponding Kaplan-Meier curves were provided with information on absolute and relative survival for these subgroups. Sources of the information were derived mostly from the Munich Tumor Registry, other sources were, for example, the website UpToDate or clinical or epidemiological studies. These epidemiological data were intended to be used as numerical 'anchors' during the (conscious) process of prognostication.
2. The brochure was introduced to the participating physicians within a teaching lecture during which a general approach to prognostication was provided (see online supplemental files). In brief, this approach assumes that prognostic accuracy can be improved by first taking an 'outside view' on a situation—for example, by using anchor numbers derived from the brochure ('actuarial' approach).[39] In subsequent steps, a more 'personalised' and multifaceted 'inside view' may follow in which relevant details of the individual patient can be used to adjust the prognostic estimate—using the principle of 'Fermisation' (referring to Enrico Fermi—see online supplemental files).

In addition, the SQ of the questionnaire was reworded in phase 2 to exclude the subjective factor of the term 'surprise'. ('*Do you think it is more likely (> 50%) that your patient will be alive in 6 months than that he will have died by that time?*')

Prior to the start of phase 3, prognostic tools were provided for the most common oncological diseases of the clinic, if they were available either as a website or as an app (see online supplemental file 3). In this phase, physicians had to provide two separate PE on the same patient—the first without knowledge of the PE made by the respective (electronic) prognostic tool and the second while being aware of the result.

In phase 3, the questionnaire was complemented by another version of the SQ—that is, the 'probabilistic SQ': In addition to the answer option 'yes/no', the SQ was now to be answered probabilistically ('*Please indicate the probability (in 10% increments from 0% to 100%) that your patient will be alive in 6 months.*').

Statistical evaluations

CPS was evaluated using the three categories 'accurate', 'neutral' and 'inaccurate'. When the real OS time was set as 100%, we set the definition of 'accurate' estimates as those that ranged from 67% to 133% of this value. 'Neutral' estimates were in the range of 34%–67% and 133%–166% and 'inaccurate' estimates were beyond these limits, that is, <34% and >166% of the actual observed survival time.

The SQ was considered 'correct' if the answer was 'yes' and the patient was alive at 6 months from the time of the

PE or the answer was 'no' and the patient had died within 6 months. In the probabilistic form of the SQ, we defined estimates as 'correct' if the patient had died at 6 months when given answers ranging from 0% to 40% or lived at 6 months when given answers ranging from 50% to 100%.

The rate of correctly answered SQ and the rate of 'accurate' CPS PE of the 'whole group' were collected as the primary endpoints of this study.

As a secondary endpoint, we aimed to describe the potential impact of a 'prognosis training' on the prognostic performance of physicians. To analyse this intervention, performance in phase 1 ('untrained') was compared with the performance after training (both phases 2 and 3 after training—'trained') and tested for statistical significance by using a χ^2 test. These analyses were performed for the 'whole group' and for the two subgroups 'residents' and 'oncologists' for each of the 3 phases.

In addition, 'residents' and 'oncologists' were tested against each other for differences in prognostic performance in each phase.

For the probabilistic form of the SQ, we tested how far the estimated survival likelihood correlates with the actual survival (χ^2 test for linear trend). The basis of this analysis is the groups of patients who had in common their respective estimated survival likelihoods—that is, '0% group', '10% group', '20% group', etc up to the '100% group'. For example, the '100% group' comprises all patients, which had been given a survival likelihood of 100% at 6 months by their respective physicians. In this probabilistic setting we also calculated the Brier score for the overall group and the subgroups (oncologists and residents).

The programme Microsoft Excel was used for descriptive statistics of the 3 phases as well as for documentation, analysis and creation of diagrams. The presentation of Kaplan-Meier curves was realised with the statistical program IBM SPSS Statistics 29.0.

Patient and public involvement

None.

RESULTS

Test persons/assessing physicians

21 physicians participated by providing a total of 2486 prognostic estimates on 748 individual patients. 18 physicians could be included, 3 were excluded because they provided <10 PE.

The whole group was subdivided into two groups. The 'residents' group was formed by rather inexperienced physicians and consisted of 11 different residents. The 'oncologists' group consisted of seven experienced board-qualified haematologists-oncologists. The professional experience of the residents averaged from entry level to 6 years of experience in internal medicine with a maximum experience in clinical oncology of 1 year. The oncologists had an experience in clinical oncology ranging from 13 to 34 years.

Table 1 Patients' characteristics

	n	in %
Total	736	100.0
Male	388	52.7
Female	348	47.3
Age (median/range)	66 (18–93)	
Pulmonary/thoracic	165	22.4
Gastrointestinal	145	19.7
Gynaecological/breast	81	11.0
Haematological	236	32.1
Urological	26	3.5
Soft tissue	11	1.5
Brain tumour	4	0.5
Cancer of Unknown Primary	15	2.0
Other	53	7.2
Localised	221	30.0
Metastatic/generalised	515	70.0
Median survival	2.5 years	

Patients

A total of 748 patients with a wide range of diseases were assessed—12 of which were excluded from the analysis because they either had no oncologic disease or no status on survival was available at the time of follow-up, so the final analysis consisted of 736 patients. The patients' characteristics can be seen in table 1. Of 736 patients, 410 (55.7%) had died by the time of the last follow-up in March 2023. The median overall survival was 2.5 years (see figure 1A).

Clinician prediction of survival

Comparison of phase 1 with phase 2 and 3

For CPS—that is, 'long-term' prognostication—1506 PE were available for evaluation. Overall, only 25.9% of them were 'accurate' according to the definition (±33% of actual observed survival). Accuracy varied over real survival as shown in figure 1B—with the best results between 8 and 10 months. Over the 3 phases, there was a numerical trend towards a slightly better accuracy from 23.6% to 28.0%—which was also observed within the two subgroups (residents and oncologists)—which however lacked statistical significance in all subgroups (see table 2).

When Kaplan-Meier curves of real and estimated overall survival were plotted against each other, the curves crossed at about 9 months for the whole group of patients. There is overestimation of OS in the 'immediate' future whereas prognosis in the more distant future tends to be underestimated. This finding is also backed by the dot plot (figure 1C) which depicts the estimated survival over real survival. Corresponding analyses of the 3 phases

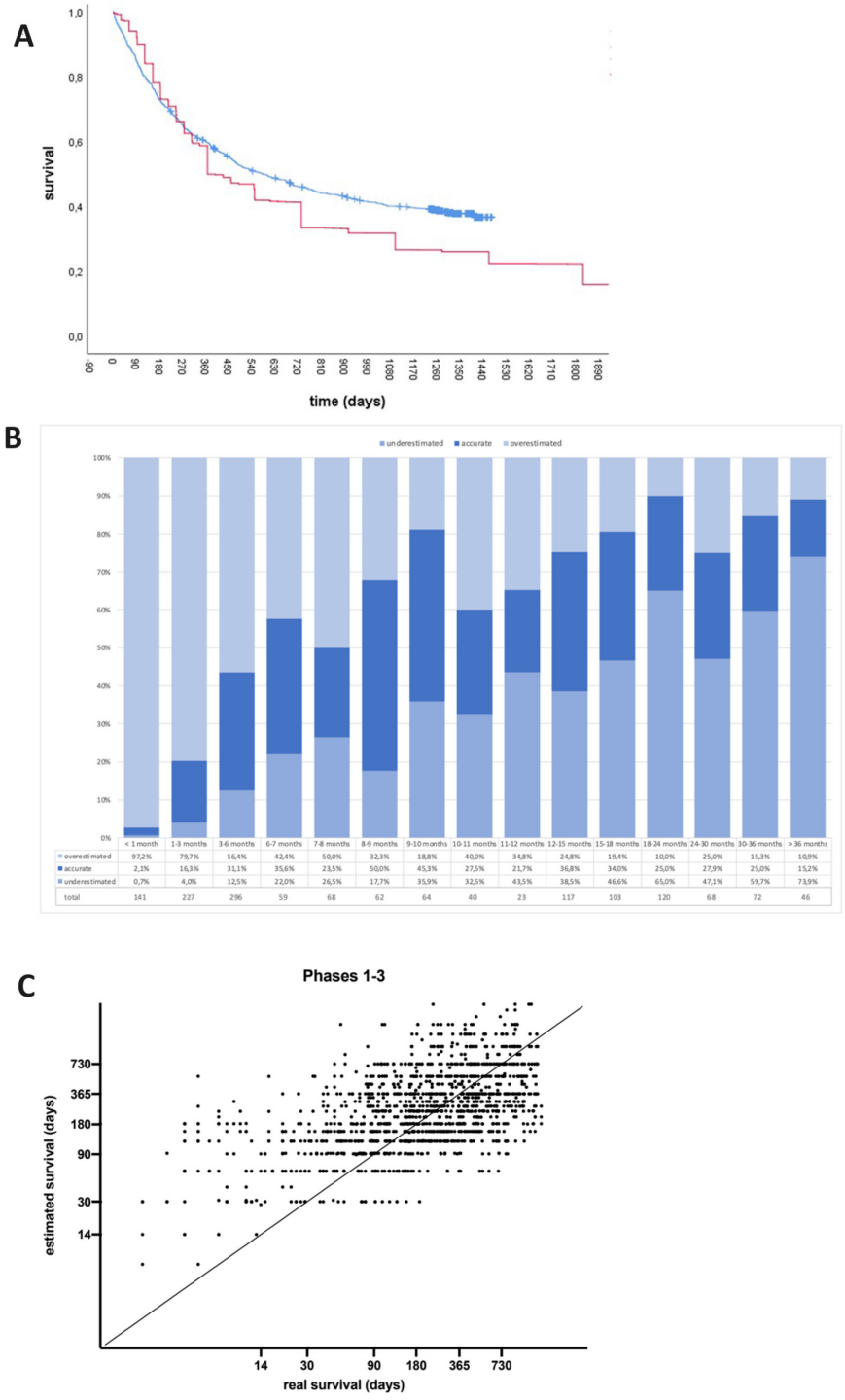


Figure 1 (A) Overall survival of patients in all phases (n=736) in blue—estimated survival (n=2427 estimates) in red. (B) Real survival (x-axis) with the annotated range (eg, 9–10 months) and the percentage of accurate prognoses (dark blue), overestimated prognoses (= high inaccurate + high moderate, light blue) and underestimated prognoses (= low inaccurate + low moderate, middle blue) in this cohort of patients. Accuracy changes with the real survival with best results between 8 and 10 months. (C) Overall survival of patients in all phases—real survival (x-axis) versus estimated survival (y-axis)—the diagonal line represents perfect prediction. Patients above diagonal are those in whom survival was overestimated; patients below line are those in whom survival was underestimated[21].

can be found in online supplemental file 4A–C. Up to a real survival of 9 months, we observed a tendency towards overestimation (671/853 (78.7%) above the diagonal of 'perfect prediction'). After 9 months, there is a tendency towards underestimation (451/653 (69.0%)) (see online supplemental file 4D).

Over time (from phases 1 to 3) the Kaplan-Meier curves tended to cross earlier—potentially indicating more

Table 2 Clinician prediction of survival phases 1+2+3

n		in % Low inaccurate (95% CI)	Low moderate (95% CI)	Accurate (95% CI)	High moderate (95% CI)	High inaccurate (95% CI)
Phase 1						
571	Total	11.6 (9.1 to 14.5)	19.8 (16.6 to 23.3)	**23.6 (20.2 to 27.3)**	8.9 (6.7 to 11.6)	36.1 (32.1 to 40.2)
153	Residents	13.1 (8.2 to 19.5)	17.6 (12.0 to 24.6)	**19.6 (13.6 to 26.8)**	10.5 (6.1 to 16.4)	39.2 (31.4 to 47.4)
418	Oncologists	11.0 (8.2 to 14.4)	20.6 (16.8 to 24.8)	**25.1 (21.0 to 29.6)**	8.4 (5.9 to 11.5)	34.9 (30.4 to 39.7)
	Residents vs Oncologists	p=*0.902759232*				
Phase 2						
492	Total	8.9 (6.6 to 11.8)	18.1 (14.8 to 21.8)	**26.6 (22.8 to 30.8)**	7.3 (5.2 to 10.0)	39.0 (34.7 to 43.5)
141	Residents	9.9 (5.5 to 16.1)	23.4 (16.7 to 31.3)	**20.6 (14.2 to 28.2)**	5.7 (2.5 to 10.9)	40.4 (32.3 to 49.0)
351	Oncologists	8.5 (5.8 to 12.0)	16.0 (12.3 to 20.2)	**29.1 (24.4 to 34.1)**	8.0 (5.4 to 11.3)	38.5 (33.3 to 43.8)
	Residents vs Oncologists	p=*0.55239201*				
Phase 3						
443	Total	5.0 (3.1 to 7.4)	19.4 (15.8 to 23.4)	**28.0 (23.9 to 32.4)**	7.2 (5.0 to 10.0)	40.4 (35.8 to 45.1)
151	Residents	4.6 (1.9 to 9.3)	19.9 (13.8 to 27.1)	**26.5 (19.6 to 34.3)**	7.3 (3.7 to 12.7)	41.7 (33.8 to 50.0)
292	Oncologists	5.1 (2.9 to 8.3)	19.2 (14.8 to 24.2)	**28.8 (23.6 to 34.3)**	7.2 (4.5 to 10.8)	39.7 (34.1 to 45.6)
	Residents vs Oncologists	p=*0.999962755*				
Total						
1506	Total	8.8 (7.4 to 10.3)	19.1 (17.2 to 21.2)	**25.9 (23.7 to 28.2)**	7.9 (6.6 to 9.4)	38.3 (35.8 to 40.8)
445	Residents	9.2 (6.7 to 12.3)	20.2 (16.6 to 24.3)	**22.2 (18.5 to 26.4)**	7.9 (5.5 to 10.8)	40.4 (35.9 to 45.2)
1061	Oncologists	8.6 (7.0 to 10.4)	18.7 (16.4 to 21.1)	**27.4 (24.8 to 30.2)**	7.9 (6.4 to 9.7)	37.4 (34.5 to 40.4)
	Residents vs Oncologists	p=*0.997692477*				
No significant differences between phase 1 vs (2+3): total p<0.2; residents p<0.6; oncologists p<0.4.						

Table 3 Surprise question option 'yes'/'no' phases 1+2+3

n		Correct % (95% CI)
Total		
2427	Total	**76.8 (75.1 to 78.5)**
571	Residents	**69.5 (65.6 to 73.3)**
1856	Oncologists	**79.1 (77.2 to 80.9)**
	Residents vs Oncologists	p<*0.0005*
Phase 1		
925	Total	**72.6 (69.7 to 75.5)**
191	Residents	**60.7 (53.4 to 67.7)**
734	Oncologists	**75.7 (72.5 to 78.8)**
	Residents vs Oncologists	p<*0.0005*
Phase 2		
712	Total	**74.0 (70.6 to 77.2)**
152	Residents	**63.2 (55.0 to 70.8)**
560	Oncologists	**77.0 (73.3 to 80.4)**
	Residents vs Oncologists	p<*0.005*
Phase 3		
790	Total	**84.3 (81.6 to 86.8)**
228	Residents	**81.1 (75.4 to 86.0)**
562	Oncologists	**85.6 (82.4 to 88.4)**
	Residents vs Oncologists	p<*0.3*

Significant differences between phase 1 vs (2+3): total p<0.001; residents p<0.01; oncologists p<0.05.

conservative estimates after training. When comparing differences between the 'estimated' CPS and the 'real' OS curves there is a visual trend towards closer curves after training (phase 2 and 3) as compared with the 'untrained' phase 1 in the 'immediate' future (see online supplemental file 4). Overestimation of survival—which was very evident in phase 1—seemed to be less pronounced after training (no test for significance).

Surprise question

Option 'yes'/'no'

In the evaluation of the SQ with the answer option 'yes'/'no', a total of 2427 answers were included. 76.8% of them were correct and 23.2% incorrect. Physicians achieved 72.6% accuracy in phase 1, 74.0% in phase 2 and 84.3% in phase 3. There was a highly significant (p<0.001) improvement in accuracy over time during the 3 phases for the whole group. Residents had an accuracy of 60.7% in phase 1, 63.2% in phase 2 and 81.1% in phase 3, which demonstrated a significant (p<0.01) improvement. The accuracy of the oncologists also increased significantly (p<0.05) from 75.7% in phase 1 to 77.0% in phase 2 to 85.6% in phase 3. In total, oncologists achieved a significantly (p<0.005) higher accuracy (79.1%) as compared with residents (69.5%). However, this was only significant in phase 1 and in phase 2, but not in phase 3 (see table 3).

Probabilistic SQ

800 answers were available for the evaluation of the probabilistic SQ in phase 3 of which 83.1% were correct. The probabilistic SQ showed a high accuracy in the extremes of the spectrum, the answer '0%' and '100%' even showed an accuracy of 100.0% with 27/27 correct SQ ('0%') and 53/53 correct SQ ('100%'). The closer the answer approached the mid-range values, the lower was the accuracy. These mid-range percentages (answers 40%–60%) only had a low accuracy of 54.5% which was significantly (p<0.0001) worse compared with the 'extreme' parts of the spectrum (0%–30% and 70%–100%) with a correct estimate in 89.9% of cases. This expression of mental indecisiveness is—of course—the equivalent of a coin flip. A detailed overview of the probabilistic SQ is shown in figure 2A. The values for the Brier score were 0.193 for the whole group, 0.191 for the oncologists and 0.199 for the residents.

An interesting finding for groups of patients was that a predicted higher likelihood of survival (eg, 'group 90%' vs 'group 60%') also correlated with a higher probability of real survival in these groups in a similar magnitude (p<0.0001) (see figure 2B). Therefore, the 'coin flip' estimate is not 'worthless' but carries information because the 'group 50%' comprises a group of patients of which roughly half will have died after 6 months.

Accuracy of prognostic tools (SQ, CPS)

During phase 3, established (electronic) prognostic tools were available and suitable for only 34.6% of patients and their respective clinical situation. When used alone for the SQ the accuracy of these prognostic tools was only 70.2%, which was significantly worse (p<0.0005) than the performance of the whole group (84.3%) and of the oncologists (85.6%). It was also numerically—if not statistically significant—worse than the performance of residents (81.1%).

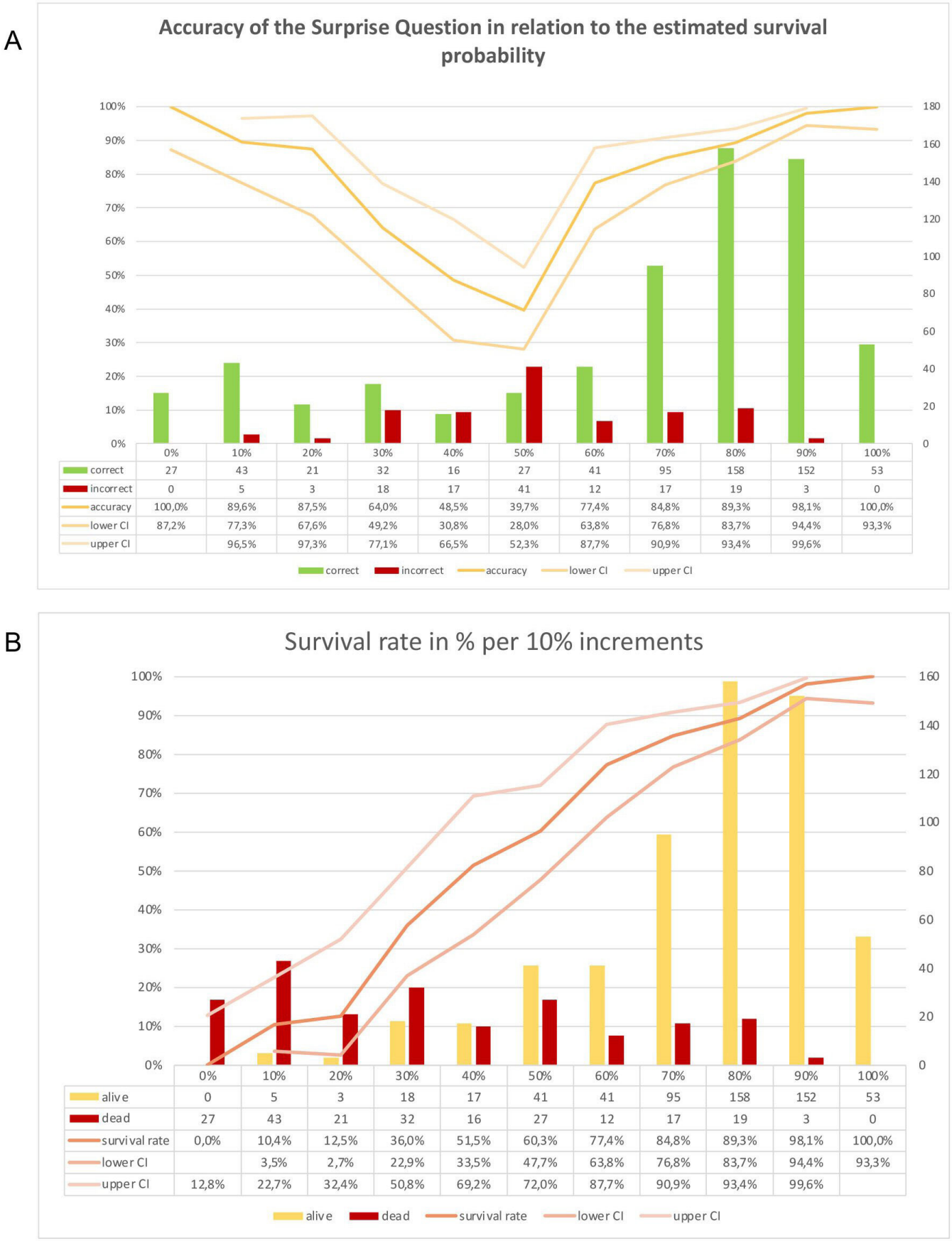
When prognostic tools were used for CPS estimates, an 'accurate' result was only achieved in 16.9% which was less than the rate of the whole group of physicians 25.9% (and of all subgroups). In a subgroup (n=443 patients) physicians made PE on CPS first without and then with the respective prognostic tool—however, accuracy was not significantly increased (28.0% to 29.4%).

DISCUSSION

This prospective study examined physicians' accuracy in prognostic estimates (PE) for oncological patients of all stages as well as a potential training effect. To do this, nearly 2500 PE on over 700 oncological patients were performed. We analysed three types of PE in our study: the (standard) SQ, the probabilistic SQ and CPS. The patients in our study had a median survival of 2.5 years, which is substantially longer than in most prior studies on prognostication in patients with cancer which have been performed on patients in very advanced stages or in a palliative setting.[2 4–16 18 19 21 25–28]

Main findings

In our study, the whole group of participating physicians achieved an accuracy of 76.8% when using the SQ. This



	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
correct	27	43	21	32	16	27	41	95	158	152	53
incorrect	0	5	3	18	17	41	12	17	19	3	0
accuracy	100,0%	89,6%	87,5%	64,0%	48,5%	39,7%	77,4%	84,8%	89,3%	98,1%	100,0%
lower CI	87,2%	77,3%	67,6%	49,2%	30,8%	28,0%	63,8%	76,8%	83,7%	94,4%	93,3%
upper CI		96,5%	97,3%	77,1%	66,5%	52,3%	87,7%	90,9%	93,4%	99,6%	

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
alive	0	5	3	18	17	41	41	95	158	152	53
dead	27	43	21	32	16	27	12	17	19	3	0
survival rate	0,0%	10,4%	12,5%	36,0%	51,5%	60,3%	77,4%	84,8%	89,3%	98,1%	100,0%
lower CI		3,5%	2,7%	22,9%	33,5%	47,7%	63,8%	76,8%	83,7%	94,4%	93,3%
upper CI	12,8%	22,7%	32,4%	50,8%	69,2%	72,0%	87,7%	90,9%	93,4%	99,6%	

Figure 2 (A) Accuracy of the surprise question (yellow line plus CI) in relation to the estimated survival probability (correct answers in green, incorrect answers in red); (B) Real survival rate in relation to the estimated survival probability (surprise question) (surviving patients in yellow, dead patients in red, survival rate=orange line plus CI).

result is fairly comparable to the summarised accuracy of 78.6% for oncological patients calculated by a systematic review in 2017.[40] It should be noted, however, that very different time intervals have been investigated so far when answering the SQ for oncological patients and that the patient cohorts differed considerably (see online supplemental file 5).[25–27] Quite evidently, the time interval chosen for the SQ has an impact on accuracy, since the longer the time interval the less certain the estimates will be and the accuracy will 'suffer'. The selection of a time interval appropriate for the respective patient population is therefore essential.

Interestingly, the prognostic accuracy of the whole group of participating physicians increased after training from an initial 72.6% to 84.3% in the last phase. This effect was especially pronounced in residents who started from a lower level (60.7% in phase 1) and then achieved 81.1% in the final phase (+20.4%). The experienced oncologists started from a significantly higher level of accuracy (75.7%) but also benefited from training (85.6% in the final phase)—however, to a less pronounced degree in absolute terms (+9.9%).

Our data indicate that experienced physicians are better prognosticators than early career physicians which has so far been a controversial issue within the literature.[8 16 21–23 41 42] Moreover, a prognosis-oriented training programme has a potential to improve the prognostic accuracy of both residents as well as of experienced oncologists. Such a focused training programme has been repeatedly recommended in the literature.[8 42 43] While some guidelines and training concepts exist in the palliative setting.[33–35] This is—to our knowledge—the first study to have implemented and systematically tested such an approach in a large group of oncological patients with a median survival over 2 years. (Our intention was to develop and test a short training programme and find out whether there is an effect. Our aim was not to specifically identify and isolate each reasoning principle within the programme and determine its incremental effectiveness.)

Wider implications

Prognostication using the SQ for a 6-month period can be considered 'short-term prognostication'. In this situation, physicians' accuracy was substantially better than chance and it was trainable. This finding contrasted with the situation of 'long-term prognostication' in which physicians were required to estimate overall survival (which most of the times exceeded 6 months—as demonstrated by a median OS of 2.5 years). The CPS was mostly 'off-target' with only 25.9% 'accurate' PE. Our results are in accordance with smaller studies which used the same definition for accuracy and demonstrated accuracy values between 20% and 35%.[6 7 11 13 15 16 18 20 21] As mentioned above, previous studies assessed very advanced cancer patients with an OS of only a few months so that survival data were available for nearly all patients.[6 7 11 13 15 16 18 21] Since only 60% of our patients had died at the time of analysis further follow-up of our patients might change the numerical value for accuracy of CPS.

CPS and SQ at 6 months are different metrics and, therefore, the absolute values of these parameters cannot be directly compared. However, CPS can be 'converted' into the format of an SQ by assuming the following: The groups 'low inaccurate', 'low moderate' and 'accurate' will not have been 'surprised' that the patient has died at his/her specific time of death. To illustrate this, imagine the following thought experiment: If we (retrospectively) know that a patient has died 450 days after discharge from our care we can pose this constellation (discharge letter from that past discharge) to a group of physicians in the form of an individual SQ: 'Would you be surprised if this patient had died within 450 days?' A physician with a CPS of '200 days' would not be 'surprised' as in our SQ definition and he would be 'right'. In contrast, a physician stating a CPS of '700 days' would be 'surprised' by the earlier demise of the patient and his estimate would have been 'wrong'. This also illustrates that the SQ is the more 'permissive' metric because it considers estimates as correct ('not surprised') that are not in the 'accurate' category of the CPS definition—that is, 'low inaccurate' and 'low moderate'.

However—even when using the more permissive 'individual SQ' definition (=consider the sum of 'low inaccurate', 'low moderate' and 'accurate' as 'not surprised' and therefore as 'correct')—the value for the CPS for the whole group is only 53.8%—hardly better than chance. Interestingly, there is no (significant) trend towards improvement by training and also there seems to be no difference between residents and experienced oncologists. These findings can be used to estimate the 'horizon' of individual clinical prognostication—which is a controversial issue in the literature[12 17 19 21 22 24 42 43]—in our study: The group of patients on which definite survival data were available (ie, had died at the time of analysis) and on which our CPS analysis was performed had a median (50%) survival of 9 months and 60% of patients had died within 1 year (data not shown). Assuming that prognostication over longer periods (than 1 year) will be even worse, we conclude that whereas individual prognostication within a time frame of 6 months by physicians is better than chance (and is trainable) this seems no longer to be the case when a time horizon of circa 1 year has been reached.

It is relevant to differentiate between prognostication for an individual patient and prognostication for a group of patients. During phase 3, when we also analysed the 'probabilistic SQ'. Its overall accuracy of 83.1% was nearly identical to that of the 'standard SQ' in that phase (84.3%) as expected. These are the numerical values that are relevant for the PE of individual patients. However, unlike the standard SQ, the probabilistic SQ also had a quantitative component ('estimate the probability of survival at 6 months in 10% increments!'). This has two consequences:

1. The forecasts in this format convey a sense of decisiveness and indecisiveness and thereby contain additional

prognostic information/certitude: This is demonstrated by the high accuracy of the extreme ends of the spectrum (0%–30% = likely to die within 6 months; 70%–100% = likely to survive the next 6 months) which reached 89.9%. This was substantially higher than in the middle ('indecisive') part of the spectrum (40%–60%) where the accuracy was only 54.4%.

2. This 'indecisive' accuracy of 54.4% in the mid-range is hardly better than chance and implies a 'coin flip' for the individual patient. However, this does not mean, that such a prognostic estimate contains no information at all and is basically worthless. In contrast, we could show that the collective prognostic estimates in 10% increments closely mirrored the de facto survival rates of groups at 6 months as demonstrated in figure 2B. For example, the 'group 50%' comprises a group of patients (all considered by their treating physician to have a 50% likelihood of survival at 6 months) of which roughly half will in fact die within 6 months.

Even though each physician performed discrete prognostic estimates on an individual patient the sum of all these forecasts nevertheless also contains prognostic meaning for groups. This 'wisdom of the clinical crowd' is also exemplified by the survival curve of our patient cohort (figure 1A) which is fitted quite well (though not perfectly) by the multitude of forecasts done on these patients by the participating physicians.

Another aspect that can be derived from these two Kaplan-Meier curves is that overestimation of prognosis tends to occur predominantly in the early phase (6–9 months) whereas underestimation tends to be predominant at the end of the first year of observation and thereafter. The two curves cross between 6 and 9 months after t_0 (= time of the prognostic estimate) with the crossing tending to be earlier in the later phases—which may have been due to more conservative forecasts after training. This finding is in accordance with Fairchild *et al* who also observed an overestimation for patients with a survival of less than 6 months and an underestimation for patients with a survival of more than 9 months.[12] We have also visualised this finding in the dot plot (figure 1C) and found a higher rate of overestimation for survival <9 months and a higher rate of underestimation for survival >9 months.

Established prognostic tools were only available for a minority of patients in their specific clinical situation (34%) and—when used alone—performed worse (accuracy 70.2%) than physicians (84.3%) in phase 3. They also did not add to prognostic accuracy of physicians. This might have been due to the fact, that often these tools are optimised for only one specific situation (eg, primary diagnosis and start of curative therapy). In these situations, some of the tools worked very well. However, most tools were not or less helpful in the many other situations when a clinical encounter takes place (eg, 3 months into second-line therapy). Also most of the time these tools convey 'collective' prognoses and place a patient into a prognostic group—similar to our results of the probabilistic SQ. This can of course be helpful, but it is a more permissive/easier task than individual prognostication.

Strengths and limitations

This is the first study that presents a structured training programme for oncologists that has a positive impact on clinical prognostication. This prospective study included a large cohort of patients with different oncological and malignant haematological diseases at all disease stages. Prospective follow-up data of almost 4 years were collected, which allowed for the assessment of longer PE. The effect of training was not demonstrated by a randomised comparison but by comparing baseline (untrained) prognostication abilities with the performance after training in the same group of physicians. The study is a single-centre study—even though performed at a large teaching hospital.

Conclusion

A short and simple training programme was able to increase the prognostic accuracy of residents as well as experienced oncologists significantly. This programme conveyed the basic principles of prognostication, provided freely available epidemiological data and taught the use of simple algorithms.[39] This teaching effect was apparent for prognostication of the intermediate future (up to 6 months) but is unlikely to be present beyond this time, especially for time horizons beyond 1 year. However, this does not come as a surprise because: 'Medicine is a science of uncertainty and an art of probability' (attributed to Sir William Osler).

Author affiliations
[1]Department of Dermatology and Allergy, Ludwig-Maximilians-Universität München, München, Germany
[2]Pepig Gabriele Thanner MS Office and project management consulting, Neutraubling, Germany
[3]Tumor Center, Centre for Quality Management and Health Services Research, University of Regensburg, Regensburg, Germany
[4]Institute for Medical Information Processing, Biometry, and Epidemiology, Pettenkofer School of Public Health, Medical Faculty, Ludwig-Maximilians-Universität München, München, Germany
[5]Department of Oncology and Hematology, Krankenhaus Barmherzige Brüder Regensburg, Regensburg, Germany

Acknowledgements We thank all members of the medical team of the department of oncology and haematology at the hospital Barmherzige Brüder who participated in this project: Berberich Veronika, Böhm Valentin, Braess Jan, Braun Bernhard, Daum Susanne, Heilmeier Bernhard, Homann Arne, Keyßner Verena, Koch Leonie, Maguire Nadia, Minderjahn Lisa, Moosmann Nicolas, Reiß Kathrin, Schenk Michael, Schlenska-Lange Anke, Seiler Isabella, Stauder Heribert, Stigler Katharina, Stosiek Christoph, Thalmeier Susanne, Wirsching Andreas.

Contributors IK, JB and GT: design and analysis of the study. IK: data collection. IK and JB: writing of the manuscript. MG: provision of survival data. IK and AC: statistical analysis. IK, GT, MG, AC and JB: critical revision of the manuscript. All authors read and approved the final manuscript. IK is the guarantor of the study.

Funding This study was completely funded by the hospital Barmherzige Brüder Regensburg.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Ethics committee of the University of Regensburg and internal review board of the hospital Barmherzige Brüder Regensburg (reference 20-2166-101). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request.

ORCID iD

Irma Kupf http://orcid.org/0009-0004-8331-8467

REFERENCES

1 Hui D, Paiva CE, Del Fabbro EG, *et al*. Prognostication in advanced cancer: update and directions for future research. *Support Care Cancer* 2019;27:1973–84.
2 Preto DD, Pantano N de P, Paiva BSR, *et al*. Clinical prediction of survival in advanced cancer outpatients: do experienced physicians and with prior patient evaluation make more accurate predictions. *J Pain Symptom Manage* 2020;59:e7–10.
3 Razvi Y, Chan S, Zhang L, *et al*. Are we better a decade later in the accuracy of survival prediction by palliative radiation Oncologists *Ann Palliat Med* 2019;8:150–8.
4 Gramling R, Gajary-Coots E, Cimino J, *et al*. Palliative care clinician overestimation of survival in advanced cancer: disparities and association with end-of-life care. *J Pain Symptom Manage* 2019;57:233–40.
5 Ermacora P, Mazzer M, Isola M, *et al*. Prognostic evaluation in palliative care: final results from a prospective cohort study. *Support Care Cancer* 2019;27:2095–102.
6 Vasista A, Stockler M, Martin A, *et al*. Accuracy and Prognostic significance of Oncologists' estimates and scenarios for survival time in advanced gastric cancer. *Oncologist* 2019;24:e1102–7.
7 Urahama N, Sono J, Yoshinaga K. Comparison of the accuracy and characteristics of the Prognostic prediction of survival of identical terminally ill cancer patients by Oncologists and palliative care physicians. *Jpn J Clin Oncol* 2018;48:695–8.
8 Tavares T, Oliveira M, Gonçalves J, *et al*. Predicting prognosis in patients with advanced cancer: A prospective study. *Palliat Med* 2018;32:413–6.
9 Farinholt P, Park M, Guo Y, *et al*. A comparison of the accuracy of clinician prediction of survival versus the palliative Prognostic index. *J Pain Symptom Manage* 2018;55:792–7.
10 Lambden J, Zhang B, Friedlander R, *et al*. Accuracy of Oncologists' life-expectancy estimates recalled by their advanced cancer patients: correlates and outcomes. *J Palliat Med* 2016;19:1296–303.
11 Amano K, Maeda I, Shimoyama S, *et al*. The accuracy of physicians' clinical predictions of survival in patients with advanced cancer. *J Pain Symptom Manage* 2015;50:139–46.
12 Fairchild A, Debenham B, Danielson B, *et al*. Comparative Multidisciplinary prediction of survival in patients with advanced cancer. *Support Care Cancer* 2014;22:611–7.
13 Perez-Cruz PE, Dos Santos R, Silva TB, *et al*. Longitudinal temporal and probabilistic prediction of survival in a cohort of patients with advanced cancer. *J Pain Symptom Manage* 2014;48:875–82.
14 Gwilliam B, Keeley V, Todd C, *et al*. Prognosticating in patients with advanced cancer--observational study comparing the accuracy of Clinicians' and patients' estimates of survival. *Ann Oncol* 2013;24:482–8.
15 Kiely BE, Martin AJ, Tattersall MHN, *et al*. The median informs the message: accuracy of individualized scenarios for survival time based on Oncologists' estimates. *J Clin Oncol* 2013;31:3565–71.
16 Hui D, Kilgore K, Nguyen L, *et al*. The accuracy of probabilistic versus temporal clinician prediction of survival for patients with advanced cancer: a preliminary report. *Oncologist* 2011;16:1642–8.
17 Gripp S, Moeller S, Bölke 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;25:3313–20.
18 Stockler MR, Tattersall MHN, Boyer MJ, *et al*. Disarming the guarded prognosis: predicting survival in newly referred patients with incurable cancer. *Br J Cancer* 2006;94:208–12.
19 Higginson IJ, Costantini M. Accuracy of prognosis estimates by four palliative care teams: a prospective cohort study. *BMC Palliat Care* 2002;1:1.
20 Llobera J, Esteva M, Rifà J, *et al*. Terminal cancer. duration and prediction of survival time. *Eur J Cancer* 2000;36:2036–43.
21 Christakis NA, Lamont EB. Extent and determinants of error in doctors' prognoses in terminally ill patients: prospective cohort study. *BMJ* 2000;320:469–72.
22 Viganò A, Dorgan M, Bruera E, *et al*. The relative accuracy of the clinical estimation of the duration of life for patients with end of life cancer. *Cancer* 1999;86:170–6.
23 Maltoni M, Nanni O, Derni S, *et al*. Clinical prediction of survival is more accurate than the Karnofsky performance status in estimating life span of terminally ill cancer patients. *Eur J Cancer* 1994;30A:764–6.
24 Bruera E, Miller MJ, Kuehn N, *et al*. Estimate of survival of patients admitted to a palliative care unit: a prospective study. *J Pain Symptom Manage* 1992;7:82–6.
25 Hamano J, Morita T, Inoue S, *et al*. Surprise questions for survival prediction in patients with advanced cancer: A multicenter prospective cohort study. *Oncologist* 2015;20:839–44.
26 Moroni M, Zocchi D, Bolognesi D, *et al*. The 'surprise' question in advanced cancer patients: A prospective study among general practitioners. *Palliat Med* 2014;28:959–64.
27 Moss AH, Lunney JR, Culp S, *et al*. "Prognostic significance of the "surprise" question in cancer patients". *J Palliat Med* 2010;13:837–40.
28 Malhotra K, Fenton JJ, Duberstein PR, *et al*. Prognostic accuracy of patients, Caregivers, and Oncologists in advanced cancer. *Cancer* 2019;125:2684–92.
29 Hui D. Prognostication of survival in patients with advanced cancer: predicting the unpredictable *Cancer Control* 2015;22:489–97.
30 Chu C, Anderson R, White N, *et al*. Prognosticating for adult patients with advanced incurable cancer: a needed Oncologist skill. *Curr Treat Options Oncol* 2020;21:5.
31 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:e0161407.
32 Cheon S, Agarwal A, Popovic M, *et al*. The accuracy of Clinicians' predictions of survival in advanced cancer: a review. *Ann Palliat Med* 2016;5:22–9.
33 Stone P, Buckle P, Dolan R, *et al*. Prognostic evaluation in patients with advanced cancer in the last months of life: ESMO clinical practice guideline. *ESMO Open* 2023;8:101195.
34 White N, Oostendorp LJ, Tomlinson C, *et al*. Online training improves medical students' ability to recognise when a person is dying: the oracles randomised controlled trial. *Palliat Med* 2020;34:134–44.
35 Bruun A, White N, Oostendorp L, *et al*. An online randomised controlled trial of Prognosticating imminent death in advanced cancer patients: Clinicians give greater weight to advice from a Prognostic algorithm than from another clinician with a different profession. *Cancer Med* 2023;12:7519–28.
36 Chang W, Chen E, Mellers B, *et al*. Developing expert political judgment: the impact of training and practice on Judgmental accuracy in geopolitical forecasting tournaments. *Judgm Decis Mak* 2016;11:509–26.
37 Tetlock PE, Mellers BA, Scoblic JP. Bringing probability judgments into policy debates via forecasting tournaments. *Science* 2017;355:481–3.
38 Kahneman D. Thinking, Fast and Slow. Toronto: Anchor Canada, 2013.
39 Meehl PE. Ovid. Clinical Versus Statistical Prediction: A Theoretical Analysis and A Review of the Evidence. Minneapolis: University of Minnesota Press, 1954. Available: http://content.apa.org/books/11281-000

40 White N, Kupeli N, Vickerstaff V, *et al*. "How accurate is the 'surprise question' at identifying patients at the end of life? A systematic review and meta-analysis". *BMC Med* 2017;15:139.
41 Maltoni M, Caraceni A, Brunelli C, *et al*. Prognostic factors in advanced cancer patients: evidence-based clinical recommendations--a study by the steering committee of the European Association for palliative care. *J Clin Oncol* 2005;23:6240–8.
42 Glare P, Virik K, Jones M, *et al*. A systematic review of physicians' survival predictions in terminally ill cancer patients. *BMJ* 2003;327:195–8.
43 Chow E, Harth T, Hruby G, *et al*. How accurate are physicians' clinical predictions of survival and the available Prognostic tools in estimating survival times in terminally ill cancer patients? A systematic review. *Clin Oncol* 2001;13:209–18.