Screening for major depressive disorder in adults with cerebral glioma: an initial validation of 3 self-report instruments

Alasdair G. Rooney, Shanne McNamara, Mairi Mackinnon, Mary Fraser, Roy Rampling, Alan Carson, and Robin Grant

Edinburgh Centre for Neuro-Oncology, Western General Hospital, Edinburgh, Scotland (A.G.R., S.M., R.G.); Beatson West of Scotland Cancer Centre, Gartnaval General Hospital, Glasgow, Scotland (M.M., M.F., R.R.); Department of Clinical Neurosciences, Western General Hospital, Edinburgh, Scotland (A.C.)

No depression screening tool is validated for use in cases of cerebral glioma. To address this, we studied the operating characteristics of the Hospital Anxiety and Depression Scale (Depression subscale) (HAD-D), the Patient Health Questionnaire–9 (PHQ-9), and the Distress Thermometer (DT) in glioma patients. We conducted a twin-center prospective observational cohort study of major depressive disorder (MDD), according to the *Diagnostic and Statistical Manual,* 4th edition, in adults with a new diagnosis of cerebral glioma receiving active management or "watchful waiting." At each of 3 interviews over a 6-month period, patients completed the screening questionnaires and received a structured clinical interview to diagnose MDD. Internal consistency, area under the receiver operating characteristics curve (AUC), sensitivity, specificity, positive predictive value, and positive likelihood ratio were calculated. A maximum of 154 patients completed the DT, 133 completed the HAD-D, and 129 completed the PHQ-9. The HAD-D and PHQ-9 showed good internal consistency ($\alpha \geq 0.77$ at all timepoints). Median AUCs were 0.931 ± 0.074 for the HAD-D and 0.915 ± 0.055 for the PHQ-9. The optimal threshold was 7+ for the HAD-D, but 8+ had similar operating characteristics. There was no consistently optimal PHQ-9 threshold, but 10+ was optimal in the largest sample. The DT was inferior to the multi-item instruments. Clinicians can screen for depression in well-functioning glioma patients using the HAD-D at the existing recommended lower threshold of 8+, or the PHQ-9 at a threshold of 10+. Due to a modest positive predictive value of either instrument, patients scoring above these thresholds need a clinical assessment to diagnose or exclude depression.

Keywords: depression, glioma, screening.

Clinical depression can be difficult to diagnose in patients with cancer and may often pass unrecognized.[1,2] It is known that the identification of depression can be improved by screening cancer patients with self-report questionnaires.[3,4] Indeed, the practice of screening for depression is supported by national and international health, cancer, and palliative care organizations.[5–7] Several depression screening tools have been validated in cancer patients.[8–11]

No depression screening instrument is validated for use specifically in patients with cerebral glioma, however. This is potentially important, first because the operating characteristics of screening measures depend partly on the sample in which they were validated,[12] and second because glioma patients are qualitatively different from other cancer patients. They have an infiltrative tumor, undergo destructive surgery, and often receive radiotherapy to the organ primarily implicated in depression.[13–15] Fatigue and cognitive dysfunction are additional possible sources of measurement error that could reduce the validity of patient self-report. The results of studies validating depression rating scales in cancer patients may not therefore generalize reliably to patients with glioma.

We aimed to conduct the first study of the validity of 3 popular patient-reported psychological rating scales in adults with glioma. We studied (i) their internal consistency and (ii) their operating characteristics compared with a structured clinical interview for major depressive disorder (MDD) as defined in the *Diagnostic and*

Statistical Manual of Mental Disorders, fourth edition (DSM-IV).

Methods

This study was part of a larger, twin-center, prospective cohort study of clinical depression in glioma patients. The setting was 2 tertiary neuro-oncology centers in Edinburgh and Glasgow (UK), together covering ~80% of the Scottish population. Full methodological details and outcomes relating to the frequency, independent clinical associations, and longitudinal course of depression have been published.[16]

Patients were eligible if they were aged ≥18 years, had a new histological diagnosis of cerebral glioma, and were fit to receive active therapy. Patients were ineligible if referred to palliative care at the point of diagnosis or if in the clinical opinion of the senior treating physician they were physically or cognitively unable to complete questionnaires. All consecutively presenting glioma patients were identified in both centers.

Participants were interviewed 3 times: during primary radiotherapy (in this sample, a median of 56 days post-primary surgery, defined as T1), then 3 months and 6 months after T1 (T2 and T3, respectively).

Variables

Demographic and treatment variables were recorded from clinical notes and are described elsewhere.[16]

For this analysis, specific predictor variables were the 3 self-report screening measures: the National Comprehensive Cancer Network's Distress Thermometer (DT)[17]; the Depression subscale of the Hospital Anxiety and Depression Scale (HAD-D)[18]; and the Patient Health Questionnaire–9 (PHQ-9).[19] All are free for general clinical use. Patients completed screening instruments at all 3 interviews. Scales were generally given in the order of DT, HAD-D, and PHQ-9. The DT was administered from the beginning of study recruitment. The HAD-D and PHQ-9 were added after the original protocol, which included a 20-min semistructured interview and a 10-min cognitive screen, and were found to be well tolerated by the first 20 participants. This was an a priori planned phased introduction of questionnaires, designed to ensure that patient fatigue would not compromise the entire study.

The DT is a single-item 11-point Likert scale constructed to look like a thermometer. Scores range from 0 = no distress to 10 = extreme distress.

The HAD-D is a 7-item depression screening questionnaire designed for use in medical populations. Each item is rated 0–3 with a maximum score of 21. Higher scores indicate greater severity of depressive symptoms over the preceding week. In general use there is a choice of 2 recommended thresholds: 8+ (for greater sensitivity) and 11+ (for greater specificity).

The PHQ-9 is a 9-item questionnaire consisting of the symptoms of MDD as currently defined by the American Psychiatric Association.[20] Each item is scored from 0 = "not at all" to 3 = "nearly every day," with a maximum score of 27. In the original validation study, which was conducted in a population of primary care outpatients, the optimal threshold was 10+.

The outcome variable was MDD. At each interview, each patient received a face-to-face Structured Clinical Interview for DSM-IV (SCID) to diagnose MDD.[21] The interviewer (A.G.R.) was a psychiatric trainee under the supervision of a consultant neuropsychiatrist (A.C.). Symptoms were counted if present—ie, no causal attributions were made. If depression was diagnosed, the patient's general practitioner and treating clinical team were informed and asked to treat the patient as they normally would. At the end of the study, patients were returned to the care of their treating team and general practitioner.

Statistical Analyses

The internal consistency of the HAD-D and PHQ-9 was examined using Cronbach's alpha.[22] Our a priori threshold of acceptable reliability was α ≥ 0.60. Item-total correlations were calculated for both scales, taking an a priori threshold of 0.40 as acceptable. For items with a low correlation, we judged pragmatically whether their removal would be likely to affect internal reliability to a clinically significant degree. Internal consistency and item-total correlations could not be examined for the single-item DT.

For each scale, the operating characteristics (sensitivity, specificity, positive predictive value [PPV], and positive likelihood ratio [LR$^+$]) were studied at each timepoint using analysis of a receiver operating characteristic (ROC) curve and classification tables. The area under the curve (AUC) was calculated to quantify the ability of each scale to discriminate between patients with and without MDD. Optimal thresholds were selected according to the best balance of operating characteristics. To estimate the likelihood that the screening measures might miss depressed patients, we also calculated the proportion of those with MDD who scored a "floor" value on each scale (defined pragmatically as a total scale score of 0, 1, or 2).

SCIDs were audio-recorded. A random sample of 10% of interviews were rescored by a consultant neuropsychiatrist (A.C.) blinded to the study diagnosis. Interrater reliability of diagnoses was calculated using Cohen's kappa.

The same researcher administered the screening questionnaires and SCIDs, initially in that order. A small number of patients with motor problems needed assistance to complete the scales, so this process could potentially have biased the SCID outcomes. To address this methodological concern, halfway through recruitment the order of interventions was reversed so that SCIDs were given first. In a post-hoc analysis, we then estimated the likelihood of expectation bias. The sample was split in two around the midpoint number of patients recruited. A chi-square analysis was performed to compare

the frequency of MDD diagnoses in the first half with that in the second half (when, with the SCIDs now being fully scored before the questionnaires, expectation bias in the primary outcome was reduced).

For reasons of power, we did not attempt to measure the sensitivity to change or test-retest reliability of the instruments.

The study was prospectively approved by the Scotland Multi-Centre Research Ethics Committee (ref 07/MRE00/55) and the local National Health Service Research and Development boards in both centers.

Results

Participants and Questionnaire Completion

Baseline clinical and demographic characteristics of participants in the parent study are presented in Table 1. The sample was broadly representative of reasonably well-functioning, newly diagnosed glioma patients.

Of 155 patients participating at T1, 154 completed the DT, 133 completed the HAD-D, and 129 completed the PHQ-9. At T2, 108 patients remained in the study, with screening questionnaire completion at 103 (DT), 91 (HAD-D), and 87 (PHQ-9). At T3, 88 patients were followed up, with questionnaire completion at 83 (DT), 80 (HAD-D), and 77 (PHQ-9). The most frequent reason for study dropout was clinical deterioration or death.

Prevalence of MDD in the Sample

At T1, 21/155 patients (13.5 $\pm$ 5.4%) had MDD; at T2 and T3 the respective frequencies were 16/108 (14.8 $\pm$ 6.7%) and 6/88 (6.8 $\pm$ 5.3%). Across all 3 timepoints a total of 32/155 individuals were diagnosed with MDD (20.6 $\pm$ 6.4%). Interrater reliability of MDD diagnosis was good ($\kappa = 0.81$, 95% confidence interval = 0.60–1.00).

HAD-D and PHQ-9 Internal Consistency/Item-Total Correlations in Glioma

For the HAD-D, across all 3 timepoints, internal consistency was good (median $\alpha = .82$). Item-total correlations for individual scale items were acceptable ($\rho = 0.42$–0.70 throughout). The HAD-D item "I feel as if I am slowed down" correlated relatively poorly with total score.

For the PHQ-9, internal consistency was good (median $\alpha = 0.85$) and item-total correlations mostly acceptable ($\rho = 0.31$–0.74 throughout). The PHQ-9 item inquiring about "Thoughts that you would be better off dead or of hurting yourself in some way" correlated least well with total score at all 3 timepoints.

All scales showed significant floor effects. At T1, these were greatest for the DT, with 81/155 (52.3%) reporting a score of 0, 1, or 2. The proportion for the HAD-D total scale score was 61/133 (45.8%). The

Table 1. Demographic characteristics of participants in the main study at T1

Variable[a]	Value
Age mean (range, SD)	54.2 (19–76, 12.3)
Patient sex	
Male	89 (57.4)
Female	66 (42.6)
Marital status	
Married	110 (71.0)
Cohabiting	19 (12.3)
Single	26 (16.8)
Glioma histology	
Glioblastoma	113 (72.9)
Other[b]	42 (27.1)
WHO glioma grade	
1–2	22 (14.2)
3–4	133 (85.8)
Hemispheric laterality	
Right	72 (46.5)
Left	72 (46.5)
Both	11 (7.1)
Tumor location	
Frontal lobe	45 (29.0)
Other[c]	110 (71.0)
Extent of resection	
Biopsy	39 (25.2)
Debulking	116 (74.8)
Radiotherapy	
Radical	121 (78.1)
Palliative[d]	18 (11.6)
None	16 (10.3)
Chemotherapy	
Temozolomide	77 (49.7)
Other[e]	9 (5.8)
None	69 (44.5)
Dexamethasone	
Yes	108 (69.7)
No	47 (30.3)
Mean dose, mg (range, SD)	2.6 (0–15, 2.7)
Antiepileptic drugs	
Yes	83 (53.5)
No	72 (46.5)
Seizures in the preceding month	
Yes	33 (21.3)
No	122 (78.7)
Karnofsky performance status	
100	25 (16.1)
90	58 (37.4)
80	44 (28.4)
70	14 (9.0)
<70	14 (9.0)
MMSE mean (range, SD)	28.2 (20–30, 1.9)

Abbreviations: WHO, World Health Organization; MMSE, Mini-Mental State Examination.
[a]Figures are n (%) except where otherwise indicated.
[b]Astrocytoma, n = 20; oligodendroglioma, n = 12; oligoastrocytoma, n = 3; pleomorphic xanthoastrocytoma, n = 1; gliosarcoma, n = 3; primitive neuroectodermal tumor, n = 2; dysembryoplastic neuroepithelial tumor, n = 1.
[c]Temporal, n = 24; parietal, n = 19; occipital, n = 7; mixed lobes/deep structures, n = 60.
[d]Palliative radiotherapy was either 30 Gy in 6 fractions (n = 12) or 40 Gy in 15 fractions (n = 6).
[e]Gliadel, n = 8; procarbazine/lomustine/vincristine, n = 1.

PHQ-9 includes somatic items and its floor effects were correspondingly lower (31/129; 24.0%).

The internal consistency and item-total correlations for the HAD-D and PHQ-9 at all 3 timepoints are shown in Table 2.

Operating Characteristics of the 3 Instruments in Glioma

For the DT, AUC was 0.88 ± 0.09 at T1, 0.90 ± 0.06 at T2, and 0.78 ± 0.20 at T3. Overall, the optimal threshold was 5+. Throughout the study, 3/32 patients with MDD reported a floor value of distress.

For the HAD-D, AUC was 0.93 ± 0.07 at T1, 0.98 ± 0.02 at T2, and 0.89 ± 0.10 at T3. Overall, the optimal threshold was 7+. However, the more traditional threshold of 8+ had similar specificity and predictive value across the study timepoints. One patient with MDD scored a floor value on the HAD-D.

For the PHQ-9, AUC was 0.92 ± 0.06 at T1, 0.94 ± 0.05 at T2, and 0.89 ± 0.11 at T3. No clearly optimal threshold was identified across the timepoints surveyed. In the largest sample (T1), providing the strongest data, the optimal threshold was 10+. No patient with MDD scored a PHQ-9 floor value at any timepoint.

Sensitivity, specificity, PPV, and LR^+ at each sampling timepoint on the 3 scales are presented in Table 3. PPV was generally poorer at T3, possibly because of the reduced frequency of MDD at this timepoint. ROC curves comparing the 3 instruments at T1 are presented in Fig. 1.

Likelihood of Expectation Bias

After excluding MDD outcome data for the first 20 patients recruited (who did not receive the HAD-D or PHQ-9), 12/58 patients were diagnosed with MDD in the first half of the sample. Following reversal of the order of interventions, 14/78 patients were diagnosed with MDD (Pearson's $\chi^2 = 0.16$, $P = .689$), suggesting no statistical evidence of expectation bias.

Discussion

Main Findings

To our knowledge this is the first examination of the validity of depression screening instruments in patients with glioma. Clinicians can use either the HAD-D or the PHQ-9 to screen for MDD in well-functioning, recently diagnosed glioma patients. Both instruments showed good internal reliability and discriminated well between patients with and without MDD. In our sample, the HAD-D displayed slightly superior and more consistent operating characteristics, including a higher PPV. The single-item DT showed inferior operating characteristics compared with the 2 multi-item questionnaires.

Limitations

An important limitation of this study was the potential for expectation bias.[23] This phenomenon could explain some of the apparent discriminatory power of the HAD-D and PHQ-9. The same researcher administered the SCIDs and screening questionnaires, when the criterion standard should ideally be independently rated. There was no statistical evidence of expectation bias, however. Additionally and perhaps more revealingly, the PHQ-9 (which lists the symptoms of MDD) performed less well than the HAD-D. These observations militate against but do not exclude the possibility of bias, and results should be interpreted cautiously.

Another limitation arises from the theoretical difficulties of confidently diagnosing depression shortly after the diagnosis of glioma. We did not make alternative diagnoses of adjustment disorder or minor depressive disorder. By interviewing patients after the start of radiotherapy, however, we exceeded the 1-month period postoperatively that is recommended to elapse before diagnosing depression in cancer patients,[24] and interrater reliability of MDD diagnoses was good. Other limitations include the lack of power to explore test-retest reliability and the potential for somatic confounding by the influence of poor concentration or fatigue on questionnaire responses. Although baseline mean Mini-Mental State Examination score was 28 with little sample variability, there is the additional possibility of measurement error arising from nonspecific cognitive effects of glioma, chemoradiotherapy, antiepileptic drugs, and/or corticosteroids, particularly at later timepoints. In line with our recruitment and sampling strategy, we suggest that findings would generalize most readily to clinically cognitively intact glioma patients in the period during and shortly after primary treatment.

Results in Context of Other Literature

Both the HAD-D and the PHQ-9 showed good internal consistency. Cronbach's alpha remained above both our a priori threshold and the more conservative threshold of 0.70 recommended by others.[25] Data are consistent with studies, conducted in varied populations, reporting internal reliabilities of 0.67–0.90 for the HAD-D[26,27] and 0.83–0.89 for the PHQ-9.[28–30] Despite theoretical concerns relating to the impact of brain cancer, these 2 screening scales were as internally consistent in adults with glioma as in other populations.

We observed, however, that the HAD-D item "I feel as if I am slowed down" performed less well in this population. Our impression was that some glioma patients endorsed this statement even when euthymic. This particular item has shown poor specificity for clinical depression in patients receiving palliative care[31] and in the context of recent myocardial infarction or stroke.[32] We hypothesize that subjective psychomotor slowing, as a consequence of glioma or its treatment and unrelated to depression, may introduce a degree of somatic confounding to the HAD-D via this item. Future studies could explore its validity in greater detail.

Table 2. Internal reliability of 2 depression screening questionnaires at 3 sampling timepoints in adults with glioma

	T1				T2				T3			
	Alpha	Mean (SD)	Item-Total Correlation	Alpha (if item deleted)	Alpha	Mean (SD)	Item-Total Correlation	Alpha (if item deleted)	Alpha	Mean (SD)	Item-Total Correlation	Alpha (if item deleted)
HAD-D	0.82				0.82				0.77			
Enjoying things as usual		0.7 (1.0)	0.69	0.77		0.7 (0.9)	0.62	0.79		0.4 (0.7)	0.54	0.73
Seeing the funny side		0.3 (0.6)	0.63	0.79		0.3 (0.6)	0.67	0.79		0.2 (0.5)	0.42	0.76
Feeling cheerful		0.4 (0.6)	0.60	0.79		0.4 (0.6)	0.67	0.79		0.3 (0.6)	0.60	0.72
Feeling slowed down		1.5 (1.0)	0.42	0.83		1.5 (1.0)	0.44	0.82		1.3 (1.1)	0.44	0.77
Losing interest in appearance		0.3 (0.6)	0.47	0.81		0.5 (0.8)	0.53	0.80		0.4 (0.7)	0.44	0.75
Looking forward to things		0.5 (0.8)	0.68	0.77		0.5 (0.8)	0.70	0.77		0.3 (0.6)	0.67	0.71
Enjoying book/radio/TV		0.4 (0.7)	0.56	0.79		0.5 (0.8)	0.44	0.82		0.3 (0.7)	0.46	0.75
PHQ-9	0.81				0.86				0.85			
Little interest or pleasure		0.6 (0.9)	0.68	0.78		0.5 (0.9)	0.63	0.84		0.4 (0.8)	0.67	0.83
Depressed mood		0.5 (0.7)	0.67	0.78		0.5 (0.8)	0.72	0.84		0.4 (0.7)	0.65	0.84
Sleep change		1.1 (1.2)	0.53	0.79		0.8 (1.1)	0.53	0.86		1.0 (1.2)	0.54	0.85
Feeling fatigued		1.4 (1.1)	0.51	0.80		1.3 (1.0)	0.69	0.84		1.1 (1.1)	0.74	0.82
Appetite change		1.0 (1.2)	0.52	0.80		0.8 (1.0)	0.58	0.85		0.6 (1.0)	0.54	0.84
Feeling bad about self		0.5 (0.9)	0.42	0.81		0.4 (0.8)	0.69	0.84		0.4 (0.7)	0.68	0.83
Trouble concentrating		1.0 (1.1)	0.56	0.79		0.5 (0.9)	0.58	0.85		0.5 (0.8)	0.63	0.83
Psychomotor changes		0.6 (0.9)	0.51	0.79		0.6 (0.9)	0.55	0.85		0.4 (0.8)	0.50	0.85
Suicidal ideas		0.1 (0.4)	0.31	0.82		0.1 (0.4)	0.42	0.86		0.1 (0.3)	0.47	0.86

T1, T2, T3 are first, second, and third sampling timepoints, respectively. Item stems are paraphrased from original scales for brevity.
For mean scores, each item had a maximum score of 3.

Table 3. Sensitivity, specificity, positive predictive values, and positive likelihood ratios for 3 depression screening tools in adults with glioma

Screening Tool and Threshold	T1				T2				T3			
	Sensitivity	Specificity	PPV	LR$^+$	Sensitivity	Specificity	PPV	LR$^+$	Sensitivity	Specificity	PPV	LR$^+$
DT^a ≥												
4	0.90	0.72	0.32	3.2	0.94	0.75	0.41	3.7	0.67	0.69	0.14	2.1
5	0.80	0.84	0.42	4.9	0.81	0.80	0.45	4.2	0.67	0.82	0.22	3.7
6	0.60	0.90	0.46	5.8	0.75	0.85	0.50	5.0	0.67	0.88	0.31	5.7
7	0.50	0.96	0.63	11.1	0.62	0.87	0.50	5.0	0.50	0.92	0.33	6.4
HAD-D^b ≥												
6	0.93	0.805	0.38	4.8	1.00	0.85	0.52	6.5	0.80	0.84	0.25	5.0
7	0.93	0.907	0.56	10.0	1.00	0.88	0.59	8.7	0.80	0.88	0.31	6.7
8	0.73	0.924	0.55	9.6	0.92	0.92	0.67	12.0	0.60	0.89	0.27	5.6
9	0.73	0.958	0.69	17.4	0.77	0.96	0.77	20.2	–^{c}	–	–	–
10	0.60	0.958	0.64	14.3	0.69	0.99	0.90	53.2	0.40	0.96	0.40	10.0
11	0.40	0.975	0.67	16.0	0.62	0.99	0.89	47.3	0.20	0.96	0.25	5.0
PHQ-9^d ≥												
8	1.00	0.68	0.29	3.1	0.93	0.84	0.52	5.7	0.83	0.80	0.26	4.2
9	0.93	0.75	0.33	3.7	0.71	0.86	0.50	5.2	0.83	0.83	0.29	4.9
10	0.80	0.86	0.43	5.7	0.71	0.89	0.56	6.5	0.67	0.83	0.25	3.9
11	0.73	0.87	0.42	5.6	0.71	0.93	0.67	10.5	0.67	0.89	0.33	5.9

T1, T2, T3 are first, second, and third sampling timepoints, respectively. LR$^+$ = likelihood ratio of the probability of screening positive when MDD is present to the probability of screening positive when MDD is absent.
Patients received a Structured Clinical Interview for DSM-IV depression at each timepoint. Boxed figures indicate the best balance of operating characteristics at each timepoint.
^{a}T1 $n = 154$, T2 $n = 103$, T3 $n = 83$.
^{b}T1 $n = 133$, T2 $n = 91$, T3 $n = 80$.
^{c}No patient scored a total of 9 on this scale at this timepoint.
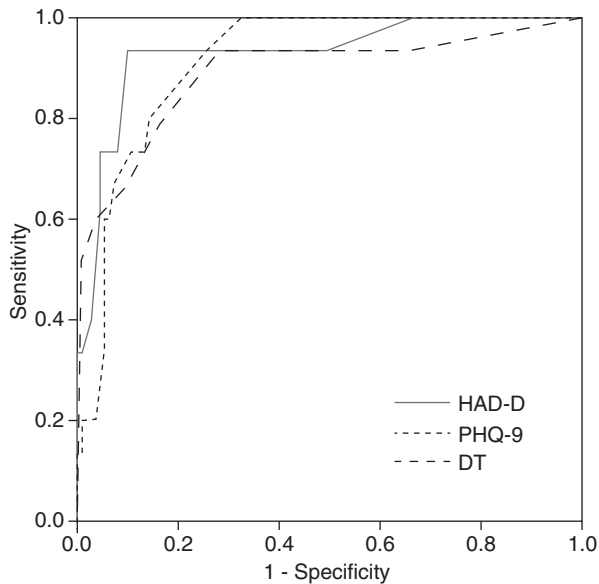^{d}T1 $n = 129$, T2 $n = 87$, T3 $n = 77$.

Fig. 1. Receiver operating characteristic curves at T1 for the HAD-D, PHQ-9, and DT, against a diagnosis of DSM-IV major depressive disorder. 95% confidence intervals are presented in the text.

Psychometrically the best HAD-D threshold was 7+. Cancer patients may require lower thresholds still than the usual lower threshold of 8+.[8]

We suggest a cautious interpretation given study limitations. Considering operating characteristics as a whole, the "tipping point" appeared to lie between scores of 6 and 9 on the HAD-D. We suggest that there is currently no good evidence to reject the lower HAD-D threshold (8+) as unsuitable in glioma patients. This threshold is familiar to clinicians and consistent with most research in cancer patients.[26] Clinicians can reasonably use the HAD-D threshold of 8+ to screen for MDD in glioma.

Head-to-head studies of the HAD-D and PHQ-9 may examine their utility as straightforward case-finding instruments or compare their ability to discriminate between different levels of depression severity. One group (including the authors of the PHQ-9) reported superiority of the PHQ-9 as a case-finding instrument in German medical outpatients.[29] Others have confirmed that both instruments show reliability, convergent/discriminant validity, and a robust factor structure but have not reported sensitivity and specificity.[28] In terms of rating depression severity, there is a small but consistent literature[31–33] suggesting that the PHQ-9 tends to overestimate and the HAD-D to underestimate severity. The scales may not therefore measure the same aspects of depression.[31]

In our sample, the HAD-D was a marginally better and more consistent case-finding instrument than the

PHQ-9. We expected the reverse because the symptoms surveyed in the PHQ-9 are identical to the syndrome of MDD. One possible reason for this result, and for its noted tendency to overestimate depressive symptom severity, is the greater likelihood of criterion confounding in the PHQ-9, which includes physical symptoms relevant to glioma, such as fatigue and appetite and sleep disturbance. By contrast, the HAD-D was designed to minimize somatic confounding even if, as we and others[34–36] have noted, an element of confounding may remain. Another possibility is that some glioma patients may find the PHQ-9 confusing. Reasonable executive function is necessary to navigate the response grid. The first item is a potential double negative ('*No loss* of interest,' paraphrasing), which some patients struggled to comprehend. The typesetting of the official version[37] also enables visually or cognitively impaired patients to accidentally record a response on the wrong line. These issues are mostly a matter of formatting and could be adapted to the needs of glioma patients.

The DT performed less well in most respects, consistent with the conclusions of a well-conducted review of single-item screening instruments for depression in cancer patients.[38] One explanation is that the DT may identify anxiety to a greater extent than it identifies depression.[39]

Conclusions

We describe an initial validation of several depression screening instruments for use in adults with glioma.

Clinicians can screen for depression in well-functioning glioma patients using the HAD-D, using the existing lower threshold of 8+, or the PHQ-9 with a threshold of 10+. The DT was inferior to the multi-item scales. Although convenient to administer, neither the HAD-D nor PHQ-9 showed sufficiently high PPV to enable clinicians to identify depression in glioma confidently through screening alone. Patients scoring high would require more detailed clinical assessment to diagnose MDD. Future research aimed at validating these scales more fully in glioma patients could examine test-retest reliability, discriminant validity, and factor structure.

References

1. Cepoiu M, McCusker J, Gole MG, Sewitch M, Belzile E, Ciampi A. Recognition of depression by non-psychiatric physicians—a systematic literature review and meta-analysis. *J Gen Intern Med*. 2007;23(1):25–36.

2. Fallowfield L, Ratcliffe D, Jenkins V, Saul J. Psychiatric morbidity and its recognition by doctors in patients with cancer. *Br J Cancer*. 2001;84(8):1011–1015.

3. Thekkumpurath P, Venkateswaran C, Kumar M, Bennet MI. Screening for psychological distress in palliative care: performance of touch screen questionnaires compared with semistructured psychiatric interview. *J Pain Symptom Manage*. 2009;38(4):520–528.

4. Vodermaier A, Linden W, Siu C. Screening for emotional distress in cancer patients: a systematic review of assessment instruments. *J Natl Cancer Inst*. 2009;101(21):1464–1488.

5. National Cancer Institute. Depression (PDQ), Health Professional Version. Available from: http://www.cancer.gov/cancertopics/pdq/supportivecare/depression/HealthProfessional. Accessed July 7, 2012.

6. National Institute for Health and Clinical Excellence. Depression in adults with a chronic physical health problem: treatment and management. London: NICE, 2009.

7. Stiefel F, Trill MD, Berney A, Olarte JMN, Razavi D. Depression in palliative care: a pragmatic report from the Expert Working Group of the European Association for Palliative Care. *Supp Care Cancer*. 2001;9(7):477–488.

8. Katz MR, Kopek N, Waldron J, Devins GM, Tomlinson G. Screening for depression in head and neck cancer. *Psychooncology*. 2004;13(4):269–280.

9. Linden W, Dahyun Y, Barroetavena MC, MacKenzie R, Doll R. Development and validation of a psychosocial screening instrument for cancer. *Health Qual Life Outcomes*. 2005;3(1):54–60.

10. Lloyd-Williams M, Dennis M, Taylor F. A prospective study to compare three depression screening tools in patients who are terminally ill. *Gen Hosp Psychiatry*. 2004;26(5):384–389.

11. Walker J, Postma K, McHugh GS, et al. Performance of the Hospital Anxiety and Depression Scale as a screening tool for major depressive disorder in cancer patients. *J Psychosom Res*. 2007;63(1):83–91.

12. Rouse SV. Using reliability generalization methods to explore measurement error: an illustration using the MMPI-2 PSY-5 scales. *J Pers Assess*. 2007;88(3):264–275.

13. Armstrong CL, Goldstein B, Cohen B, Mi-Yeoung J, Tallent EM. Clinical predictors of depression in patients with low-grade brain tumors: consideration of a neurologic versus a psychogenic model. *J Clin Psychol Med Settings*. 2002;9(2):97–107.

14. Arnold SD, Forman LM, Brigidi BD, et al. Evaluation and characterization of generalized anxiety and depression in patients with primary brain tumors. *Neurooncol*. 2008;10(2):171–181.

15. Weitzner MA. Psychosocial and neuropsychiatric aspects of patients with primary brain tumors. *Cancer Invest*. 1999;17(4):285–291.

16. Rooney A, McNamara S, Mackinnon M, et al. The frequency, clinical associations and longitudinal course of Major Depressive Disorder in adults with cerebral glioma. *J Clin Oncol*. 2011;29:4307–4312.

17. The National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology: Distress Management. Available from: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp. Accessed July 7, 2012.

18. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression scale. *Acta Psychiatr Scand*. 1983;67(6):361–370.

19. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606–613.

20. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Text Revision. 4th ed. Washington, DC: APA, 2000.

21. First MB, Gibbon M, Spitzer RL, Williams JBW. User's Guide for the SCID-I, Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version. New York: New York State Psychiatric Institute, 1996.

22. Cortina JM. What is coefficient alpha? An examination of theory and application. *J Appl Psychol*. 1993;78(1):98–104.

23. Greenhalgh T. How to read a paper: papers that report diagnostic or screening tests. *BMJ*. 1997;315(7107):540–543.

24. Satin JR, Linden W, Phillips MJ. Depression as a predictor of disease progression and mortality in cancer patients. *Cancer*. 2009;115(22):5349–5361.

25. Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *BMJ*. 1997;314:572.

26. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res*. 2002;52(2):69–77.

27. Mykletun A, Stordal E, Dahl AA. Hospital Anxiety and Depression (HAD) scale: factor structure, item analyses and internal consistency in a large population. *Br J Psychiatry*. 2001;179(6):540–544.

28. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *Br J Gen Pract*. 2008;58(546):32–36.

29. Lowe B, Spitzer RL, Grafe K, et al. Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord*. 2010;78(2):131–140.

30. Martin A, Rief W, Klaiberg A, Braehler E. Validity of the Brief Patient Health Questionnaire Mood Scale (PHQ-9) in the general population. *Gen Hosp Psychiatry*. 2006;28(1):71–77.

31. Cameron IM, Cardy A, Crawford JR, du Toit SW, Hay S, Lawton K, et al. Measuring depression severity in general practice: discriminatory performance of the PHQ-9, HADS-D, and BDI-II. *Br J Gen Pract*. 2011;e419–e426.

32. Reddy P, Philpot B, Ford D, Dunbar JA. Identification of depression in diabetes: the efficacy of the PHQ-9 and HADS-D. *Br J Gen Pract*. 2010;e239–e245.

33. Hansson M, Chotai J, Nordstom A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. *Br J Gen Pract*. 2009;e283–e288.

34. Lloyd-Williams M, Friedman T, Rudd N. An analysis of the validity of the Hospital Anxiety and Depression Scale as a screening tool in patients with advanced metastatic cancer. *J Pain Symptom Manage*. 2001;22(6):990–996.

35. Johnston M, Pollard B, Hennessey P. Construct validation of the Hospital Anxiety and Depression Scale with clinical populations. *J Psychosom Res*. 2000;48(6):579–584.

36. Natusch D. Criterion contamination when using the Hospital Anxiety and Depression Scale. *Anaesth*. 2006;61(6):609–610.

37. The Patient Health Questionnaire–9. Available at: http://www.phqscreeners.com/pdfs/02_PHQ-9/English.pdf. Accessed July 7, 2012.

38. Mitchell AJ. Pooled results from 38 analyses of the accuracy of distress thermometer and other ultra-short methods of detecting cancer-related mood disorders. *J Clin Oncol*. 2007;25(29):4670–4681.

39. Mitchell AJ, Baker-Glenn EA, Granger L, Symonds P. Can the Distress Thermometer be improved by additional mood domains? Part 1. Initial validation of the Emotion Thermometers tool. *Psychooncology*. 2010;19(2):125–133.