Chapter 2

Systematic review of brain metastases prognostic indices

Rodrigues G\(^1,2\), Bauman G\(^1\), Palma D\(^1\), Louie AV\(^1\), Mocanu J\(^1\), Senan S\(^3\), Lagerwaard F\(^3\)


2. Department of Epidemiology and Biostatistics, University of Western Ontario, London, ON, Canada, N6A3K7.

3. Department of Radiation Oncology, VU University Medical Center, de Boelelaan 1117, 1081 HV, Amsterdam, The Netherlands.

Practical Radiation Oncology 2013;3:101-106
ABSTRACT

Purpose
A variety of prognostic indices for patients with brain metastases have been published in the literature, to guide clinical decision-making and clinical trial stratification. The purpose of this investigation is to perform a systematic review of all primary and validation reports of such prognostic systems. An assessment of index operating characteristics and misclassification rates was performed to assist in highlighting the advantages and disadvantages of competing systems.

Methods and Materials
A systematic review of the English language literature regarding primary and validation brain metastases prognostic indices was performed according to PRISMA guidelines. Clinical, treatment, statistical, and prognostic index classification details were abstracted and organized into tables. Receiver operator characteristic (ROC) curves were created from available Kaplan-Meier curves using a novel digitization procedure. From these curves, various operating characteristics such as positive predictive value (PPV), negative predictive value (NPV), accuracy (ACC), likelihood ratio (LR), and area under the curve (AUC) were calculated. Additionally, the major misclassification rate (MMR), defined as good or poor risk patients misclassified into the opposite group, was calculated for all available ROC curves.

Results
A total of nine prognostic systems have been published in the medical literature. In terms of the poor prognostic group, observed ranges for PPV (0.25-0.72), NPV (0.72-0.97), ACC (0.57-0.95), LR (1.54-16.4), AUC (0.64-0.90), and MMR (0.02-0.39). Similarly, ranges of PPV (0.52-0.96), NPV (0.31-0.77), ACC (0.41-0.74), LR (1.69-20), AUC (0.64-0.89), MMR (0.00-0.19) were observed for the good prognostic group.

Conclusions
Operating characteristic and major misclassification analyses of all available prognostic index information demonstrated a range of results. As the ideal prognostic index has not yet been defined, further research into alternative approaches is warranted. Information contained within this report can serve as a benchmark for future investigations of existing and proposed prognostic indices.
INTRODUCTION

Prognostic indices are commonly utilized in a variety of cancer scenarios in which knowledge regarding expected patient prognosis may impact medical decision-making and patient treatment selection. Additionally, these indices are routinely used to define criteria for clinical trial eligibility and stratification of randomization in order to support the proper design of clinical trials. Well-known examples of primary cancer prognostic indices that are commonly employed in the context of clinical care include: breast cancer, prostate cancer, lymphomas and germ cell tumours. Additionally, several prognostic indices have been described in regards to patients with brain metastases, which have been utilized both in the conduct of clinical trials and patient care [1].

Brain metastases are a common manifestation of cancer, and an important cause of patient morbidity and mortality. Whole brain radiotherapy (WBRT) is commonly used for palliation, with more aggressive therapy reserved for good prognosis patient populations. In favourable-prognosis patients, neurosurgical and stereotactic radiosurgical (SRS) approaches can improve overall survival, yet these overall survival benefits may not extend to poor-prognosis groups [2-3]. Therefore, a clinical imperative of appropriate patient selection into good and poor prognostic groups exists in order to optimize the therapeutic ratio between survival and local control versus treatment toxicity, and to minimize risks of over- and under-treatment. To address this need, various brain metastases prognostic indices have been described in the literature to support medical decision-making and clinical trials in this patient population [4-13].

The objective of this systematic review is to assess published prognostic indices for brain metastases in terms of patient population composition, statistical methodology, prognostic modeling procedures and group definitions, and prognostic classification distribution. The ability of the indices to classify patients into favourable and non-favourable prognostic groups will be assessed by calculation of various operating characteristics from the available medical literature. A comparison of all existing prognostic indices will be performed to assess index advantages and disadvantages as well as statistical validity.

METHODS AND MATERIALS

Search Strategy
A systematic review of the literature was performed according to PRISMA systematic review
guidelines (www.prisma-statement.org). The English-language MEDLINE literature was searched for
relevant papers between 1966 to December 31, 2011 that met the following inclusion criteria:

1. Primary reports of brain metastases prognostic classification systems and/or prognostic
   indices.

2. Reports providing validation information for these classification systems and/or prognostic
   indices.

The following inclusion criteria were applied to identify the final manuscripts to be used for data
abstraction and analysis:

1. Studies reporting on generic brain metastases populations, without focusing on specific
   primary cancer or brain metastases populations (e.g. lung primaries only, solitary
   metastases).

2. Studies validating previously published prognostic indices must include at least 200
   patients, present Kaplan-Meier curves, and document the relative proportions of patients in
   each class. These criteria were applied to ensure that the data would assist in the qualitative
   analysis of prognostic index operating characteristics.

Data Abstraction

The following information was abstracted from all primary and validation reports: primary author,
reference, year of publication, number of patients, patient population demographics, treatment, statistical
methodology, prognostic factors assessed, final model prognostic factors, prognostic index scoring
system, survival and patient proportion within each class, and study conclusions.

Operating Characteristic Analysis

For manuscripts to be included in the operating characteristic analysis, all classification groups
need to be represented in the manuscript. One paper was excluded from the quantitative operating
characteristic analysis due to the lack of any patients in the poor prognostic group [14]. In order to create
receiver operating characteristic curves from eligible primary and validation manuscript, relevant Kaplan-
Meier curves depicting survival estimates per classification group were imported into Microsoft Excel (Microsoft Inc., Redmond WA, USA) using an in-house digitization procedure. Death events were estimated by the product of relative survival at a specific time-point with the total number of patients within the class being investigated.

From the tabular death information, receiver-operating curves for early death (defined as death within two months) and late survival (defined as survival after six months) were constructed for the high-risk (poor prognostic group - PPG) and low-risk (good prognostic group - GPG) prognostic classifications, respectively as described by Nieder and Metha [1]. Reported 2-month high-risk and 6-month low-risk operating characteristics included positive predictive value (PPV, proportion of patients in a risk group developing the event of interest – death for high-risk, alive for low-risk), negative predictive value (NPV, proportion of patients not in a risk group that do not develop the event of interest), accuracy (ACC, proportion of all patients correctly classified), likelihood ratio (LR, related to the value of a classification system to change the underlying probability that an event of interest will occur), and area under the curve (AUC, estimate of predictive power of the index at all time-points). A major misclassification rate (MMR) was calculated for both 2-month high-risk and 6-month low-risk groups. This was defined as the proportion of patients in the high-risk group alive after 6 months or the proportion of patients in the low-risk group that died within 2 months.

A summary of the statistical properties, advantages, and disadvantages of all classification systems or prognostic indices was performed. This was based on the qualitative data abstraction information and the quantitative receiver-operator curve analysis of all eligible primary and validation reports.

RESULTS

Literature Search

A total of 396 records were initially assessed for systematic review eligibility with a total of 59 articles undergoing full text review (Figure 1). A total of eighteen studies met all review criteria with a total of nine prognostic systems being reported [4-13]. Ten reports were related to the initial description of the prognostic index with two overlapping reports for the Disease-Specific Graded Prognostic Index (DS-GPA) [11,12]. Seven unique validation studies were identified for inclusion in the systematic review database [14-20]. One primary report on the original Graded Prognostic Index (GPA) contained validation information on three other indices [8]. A total of 13/17 reports contained Kaplan-Meier and
classification distribution information that allowed for operating characteristic calculations, and these 13 studies were used for analysis.

**Prognostic Indices**

The nine unique prognostic systems were identified from the medical literature were: the Radiation Therapy Oncology Group Recursive Partitioning Analysis (RTOG RPA), the GPA and DS-GPA, the Rotterdam System (ROTTERDAM), the Score Index of Radiosurgery (SIR), the Basic Score for Brain Metastases (BSBM), the Golden Grading System (GGS), and two Rades classifications (RADES I and II). Various parameters (e.g. sample size, source data, details of radiation, and statistical
Methodology related to these primary databases are detailed in Table 1. Data from a total of 12247 patients (some duplicated between RTOG RPA and GPA analyses) have been utilized to generate these nine prognostic indices with individual reports ranging from 65 to 4259 patients. A heterogeneous patient and treatment population is noted. Multivariable analysis was utilized in 7/9 reports with RPA techniques used for the RTOG RPA and GGS. The number of factors that were initially assessed ranged from 5 to 12, with a total of 3 to 6 factors populating the final multivariable models. Individual factors commonly utilized in the final models included performance status (9/9), extracranial metastases (9/9), age (7/9), primary controlled (4/9), and number of brain metastases (4/9) as summarized in Table 2.

<table>
<thead>
<tr>
<th>PS</th>
<th>Author [REF]</th>
<th>Source Data</th>
<th>N</th>
<th>Statistical Methods</th>
<th>Factors Assessed</th>
<th>Final Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTOG RPA</td>
<td>Gaspar [4], 1997</td>
<td>RTOG 7916, 8522, and 8965</td>
<td>1200</td>
<td>RPA with KM/WM</td>
<td>Age, KPS, neurological function/signs/symptoms, primary pathology, PC status, ECM, number/interv BM, BM characteristics, prior surgery, radiation dose and response</td>
<td>KPS, PC, Age, and ECM</td>
</tr>
<tr>
<td>ROTTERDAM</td>
<td>Lagenwaard [5], 1999</td>
<td>Daniel den Hoed Cancer Center</td>
<td>1202</td>
<td>MVA with KM/LR</td>
<td>Age, sex, ECOG PS, number/distribution/interv BM, primary site, pathology, systemic met, LDH, response to steroids, treatment</td>
<td>ECOG PS, ECM, PC, steroid response</td>
</tr>
<tr>
<td>SIR</td>
<td>Weitman [6], 2000</td>
<td>Hospital Israelita Albert Einstein</td>
<td>65</td>
<td>MVA with KM/LR</td>
<td>Age, KPS, ECM, maximum lesion volume, number/site of lesions, use of WBRT</td>
<td>KPS, PC, Age, ECM, number BM, volume BM</td>
</tr>
<tr>
<td>BS-BM</td>
<td>Lorenzoni [7], 2004</td>
<td>Gamma Knife Center, Brussels</td>
<td>110</td>
<td>MVA with KM/C</td>
<td>Age, sex, KPS, PC status and site, ECM, BM number/volume, WBRT utilization</td>
<td>KPS, PC, ECM</td>
</tr>
<tr>
<td>GPA</td>
<td>Spendulo [8], 2009</td>
<td>RTOG 7916, 8529, 8905, 9104, and 9509</td>
<td>1000</td>
<td>MVA with KM/LR</td>
<td>Age, sex, KPS, histology, BM interval, ECM, BM number</td>
<td>Age, KPS, number BM, ECM</td>
</tr>
<tr>
<td>RADES I</td>
<td>Rades [9], 2009</td>
<td>Multi-institutional</td>
<td>1085</td>
<td>MVA with KM/RC</td>
<td>Age, KPS, sex, WBRT schedule, primary site, number BM, ECM, RPA, interval to BM</td>
<td>Age, KPS, ECM, interval to BM</td>
</tr>
<tr>
<td>GGS</td>
<td>Golden [10], 2008</td>
<td>University of California at San Francisco, 2000</td>
<td>479</td>
<td>RPA with KM/C</td>
<td>Age, KPS, PC, ECM, number BM</td>
<td>Age, KPS, ECM</td>
</tr>
<tr>
<td>DS-GPA</td>
<td>Spendulo [11], 2010 and 2012</td>
<td>Multi-institutional</td>
<td>4259</td>
<td>MVA with KM/RC</td>
<td>GPA components, BM treatment, primary diagnosis</td>
<td>GPA components and primary diagnosis</td>
</tr>
<tr>
<td>RADES II</td>
<td>Rades [13], 2011</td>
<td>Multi-institutional</td>
<td>179</td>
<td>MVA with KM/RC</td>
<td>Age, sex, KPS, primary site, number BM, ECM, interval BM</td>
<td>Age, KPS, ECM, interval to BM, number BM</td>
</tr>
</tbody>
</table>

PS = Prognostic Score; REF = Reference; N = Number of Patients; RTOG = Radiation Therapy Oncology Group; RPA = Recursive Partitioning Analysis; PC = Primary Controlled; KPS = Karnofsky Performance Status; KM = Kaplan-Meier; W = Non-parametric Wilcoxon statistic; ECM = Extracranial Metastases; BM = Brain Metastases; MVA = Multivariable Analysis; LR = Log-Rank Statistic; ECOG = Eastern Cooperative Oncology Group; LDH = Lactate Dehydrogenase; PS = Performance Status; SIR = Score Index for Radiosurgery in BM; WBRT = Whole Brain Radiotherapy; SRS = Stereotactic Radiosurgery; BS-BM = Basic Score for Brain Metastases; C = Cox Proportional Hazard Analysis; GFA = Graded Prognostic Index; DS = Disease-Specific; GGS = Golden Grading System.
Four of the nine systems (RPA, ROTTERDAM, SIR, GGS) divided patients into three prognostic groups while the other five divided patients into four groups. Classification procedures and definitions, proportions of patients in each prognostic group, and the medical survival of each prognostic group are summarized in Table 3. A total of eight reports (n=4214 patients) provided prognostic index validation information for the RTOG RPA (n=7 studies), BSBM (n=2 studies), GPA (n=2 studies), and SIR (n=1 study), which is summarized in Table 4. All validation reports have generally concluded that these systems have utility; however, the primary PGA report which contained a comparative analysis against the RTOG RPA, SIR, and BSBM concluded that the BSBM system may be inferior to the RPA/GPA.
The RTOG RPA is the most extensively utilized and investigated prognostic system of those available; however, misclassifications of patients between favourable and unfavourable groups have been previously identified [1]. The initial validation is based on 1200 patients and did provide prognostic groups of reasonable size [4]. Follow-up validation studies using a variety of patient populations and statistical methodologies have continued to demonstrate the utility of the system [8, 14-18].
Two new RTOG systems have been proposed (GPA and DS-GPA) to potentially further refine prognostic power and ability [8, 11-12]. The GPA utilized a four-group paradigm and included SRS patient populations (RTOG RPA was primarily WBRT patient populations). A limitation of the GPA system is that the favourable group has been observed to comprise a small group on both the primary and validation analyses. The DS-GPA is primarily based on a large multi-institutional SRS/WBRT patient population and has integrated a diagnosis specific scoring system to better adjust for observed differences in outcome. This system has not yet undergone an independent validation.

Another reported prognostic index is the Rotterdam system that is based on single institutional data from primarily a WBRT population. This system is unique in that it utilizes information on steroid response; however, this index has not been subsequently validated and also includes relatively complex definitions of primary and extracranial metastasis control. The GGS three-group system is based on single-institution data with the advantage of a simple scoring system; however, no eligible validation reports have been published. Both the SIR and BSBM systems are simple in nature but based on a small single institution primary datasets; however, they have been subsequently studied with a larger validation databases [8, 20]. Rades et al. have published two widely cited prognostic models and systems, which are based on large multi-institutional data containing mainly WBRT data [9,13]. These two systems have balanced groups; however, no independent validation reports are available.

**Operating Characteristic Analysis**

Operating characteristics for prediction of death at 2 months (i.e. PPG) and survival at 6 months (i.e. GPG) are summarized in Tables 5 and 6, respectively. In terms of the PPG analyses, observed ranges for PPV (0.25-0.72), NPV (0.72-0.97), ACC (0.57-0.95), LR (1.54-16.4), and AUC (0.64-0.90) are detailed in Table 5. Similarly, observed ranges for PPV (0.52-0.96), NPV (0.31-0.77), ACC (0.41-0.74), LR (1.69-20), and AUC (0.64-0.89) related to the GPG analyses are detailed in Table 6.
In terms of the PPG analysis, the MMR ranged from 2 to 39% with 1/1 GGS (39%), 1/3 BSBM (26%), 1/3 GPA (31%), and 3/6 RPA (26%, 29%, 30%) analyses demonstrating PPG MMR greater than 25%. The lowest PPG MMR was associated with the ROTTERDAM and RADES I systems. The MMR associated with the GPG analyses are generally lower and range from 0 to 19% with two BSBM (15%, 19%) and two GPA (18%) analyses demonstrating the highest levels of major misclassification (i.e. GPG
patients dying within two months). Three of six RTOG RPA, 2/2 SIR, 1/3 BSBM, 1/3 GPA, and 1/1 RADES I/II/GGS analyses demonstrated GPG MMR of 10% or less.

DISCUSSION

Nine unique systems composed of various combinations of nine individual prognostic factors have been investigated in the medical literature. Although use of performance status and extracranial metastases status are common to all systems, other factors (e.g. brain metastases volume, interval to brain metastases diagnosis, and steroid response) were not consistently utilized. Significant heterogeneity in terms of patient populations, treatment approaches, statistical methodologies, factors assessed and number of prognostic groups was observed. No reports in the literature directly addressed the concepts of operating characteristics or misclassification rates. This investigation has been able to quantify the operating characteristics for GPG and PPG prognostication. Generally, a wide range of prognostic ability was observed depending on the prognostic index, patient population being studied, and the group (GPG/PPG) being assessed.

Investigations by Sperduto et al. [8] and Villa et al. [20] have provided data that can facilitate head-to-head comparisons between different prognostic indices. The Sperduto study primarily assessing the GPA index also provided information on RTOG RPA, SIR, and BSBM. In terms of the PPG operating characteristics and MMR, no major differences between systems can be observed. However, for the six-month GPG analysis, it appears that the GPA system may have advantages in terms of PPV, LR, AUC, and MMR. The Villa study assessed the GPA index against the RTOG RPA and the BSBM systems. No striking differences were seen relating to operating characteristics and MMR for both the PPG and GPG analyses; however, a slight advantage of RTOG RPA and GPA over BSBM in terms of GPG LR and AUC can be observed.

As summarized in Table 7, the observed statistical validation procedures, advantages, and disadvantages of all nine prognostic indices studied in this report must be considered. Desirable qualities associated with both the initial and subsequent validation reports include: large datasets (RTOG RPA, GPA, DS-GPA, RADES I/II, and ROTTERDAM), sufficient validation (RTOG RPA, SIR, GPA), extensive clinical utilization (RTOG RPA), balanced group proportions (RADES I/II and DS-GPA), and simplicity (RTOG RPA, GGS, BSBM). Non-desirable qualities included low PPG PPV (RTOG RPA, SIR, BSBM, GPA, RADES I), small GPG groups of less than 10% (ROTTERDAM, SIR, GGS, GPA), operating characteristics not able to abstracted from the report (DS-GPA), inferior MMR for either GPG/PPG (GGS, RTOG RPA, BSBM, GPA), and no subsequent validation reports (ROTTERDAM,
RADES I/II, DS-GPA, and GGS). Given that the RTOG RPA system contains multiple desirable qualities with no other system clearly demonstrating statistical superiority, continued use of the RTOG RPA system for clinical decision-making and clinical trial design is recommended. The ideal system should have a low PPG MMR in order to ensure that patients that may benefit from aggressive therapy have the opportunity to receive such therapy. Unfortunately, the two systems with the lowest PPG MMR (ROTTERDAM and RADES I) have not been subjected to subsequent validation. Additionally, it is important to note that the two systems with the highest reported accuracy (GGS for PPG and RADES II for GPG) have not been subjected to independent validation to confirm these operating characteristic levels.

<p>| Table 7: Advantages and Disadvantages of Brain Metastases Prognostic Indices |
|------------------------------------------|------------------|------------------|------------------|</p>
<table>
<thead>
<tr>
<th><strong>Prognostic Index</strong></th>
<th><strong>Statistical Validation</strong></th>
<th><strong>Advantages</strong></th>
<th><strong>Disadvantages</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>RTOG RPA</td>
<td>large prospectively collected RTOG clinical trial primary dataset, multiple validation investigations</td>
<td>well-known and described classification system, frequently utilized in clinical care and clinical trials</td>
<td>moderate OC with low PPG PPV, some validation studies suggest high PPG MMR</td>
</tr>
<tr>
<td>ROTTERDAM</td>
<td>large primary dataset, no validation investigations*</td>
<td>good OR for both PPG and GPG classifications, high PPG PPV, low PPG MMR</td>
<td>small GPG (&lt;10%) in primary validation, use of subjective steroid response, moderate GPG PPV</td>
</tr>
<tr>
<td>SIR</td>
<td>very small primary dataset, large validation dataset confirming OC</td>
<td>derived from whole brain plus SRS population</td>
<td>small GPG (10%) in primary validation, moderate OC with very low PPG PPV</td>
</tr>
<tr>
<td>BS-BM</td>
<td>small primary dataset, validation datasets do not confirm high OC seen in primary report</td>
<td>derived from various SRS populations, simple scoring system</td>
<td>small (3%) PPG group with low PPV, high GPG MMR</td>
</tr>
<tr>
<td>GPA</td>
<td>very large prospective RTOG clinical trial primary dataset, moderate sized validation cohorts</td>
<td>derived from WBRT and SRS patient populations, good GPG OC</td>
<td>small GPG cohorts in primary and validation datasets, low PPG PPV, high GPG MMR</td>
</tr>
<tr>
<td>RADES I</td>
<td>large primary dataset, no validation investigations*</td>
<td>balanced proportions of patients in each of four groups, overall good OC and PPG MMR</td>
<td>low PPG PPV</td>
</tr>
<tr>
<td>GGS</td>
<td>moderate size primary dataset, no validation investigations*</td>
<td>straightforward index construction, overall good OC</td>
<td>small (3%) PPG group with low PPV, high PPG MMR</td>
</tr>
<tr>
<td>DS-GPA</td>
<td>large primary dataset, no validation investigations*</td>
<td>disease-specific scoring integrated into index, balanced proportions of patients in each of four groups</td>
<td>complex index construction, OC could not be calculated from published reports</td>
</tr>
<tr>
<td>RADES II</td>
<td>large primary dataset, no validation investigations*</td>
<td>good OC, well balanced groups</td>
<td>patient population not well reported in manuscript</td>
</tr>
</tbody>
</table>

* no validation studies that meet systematic review eligibility criteria (see methods section)

PPV = Positive Predictive Value; PPG = Poor Prognostic Group; GPG = Good Prognostic Group; OR = Operating Characteristics; SRS = Stereotactic Radiosurgery; MMR = Major Misclassification Rate

This investigation has several important limitations including the fact that highly selected patient populations (e.g. solitary brain metastases or breast cancer populations) were not investigated as part of this systematic review. Operating characteristics and MMR abstracted from the reports were not from individual patient data points but were instead an estimate based on published Kaplan-Meier curves. Additionally, smaller validation studies were excluded from this systematic review and quantitative analysis but may provide additional information. Many of these studies have been included in a previously published general review of the topic [1].

35
Future publications of either existing or proposed prognostic indices should provide the following information in order to provide the medical community adequate information regarding the clinical and statistical properties of the index:

1. A detailed account of patient, tumour, and treatment factors including patient age, performance status, extracranial metastases, primary site and status, number and volume of brain metastases, interval to diagnosis, steroid response as well as radiotherapy details (volume, dose, fraction size, technique).

2. Alternative prognostic index distribution should be reported along side those being investigated in the report.

3. A detailed account of statistical procedures utilized to create or validate the prognostic index.

4. Ideally, both testing and validation procedures should be reported within the same manuscript for future indices.

5. Kaplan-Meier curves and receiver operator curves for existing and proposed indices should be presented.

6. Calculation of operating characteristics such as PPV, NPV, ACC, LR, and AUC as well as the MMR for GPG and PPG should be performed.

7. Where feasible, direct comparison of operating characteristics and MMR between existing and proposed systems should be performed.

Based on this analysis, future investigations in this area can include: creation of a rigorous definition of early death and prolonged survival (i.e. further refinement of the 2 and 6 months used for this investigation) based on available survival data from existing databases, optimization of a novel prognostic system based on artificial neural network techniques, a single institutional cross-index comparison analysis, an international database collaboration to improve statistical power for cross-index comparisons, and ultimately the generation of evidence-based guidelines to assist practitioners utilize available prognostic indices to assist in patient care decision-making.
CONCLUSIONS

Multiple primary and validation investigations into nine brain metastases prognostic indices have been described in the medical literature. Cross-validation between prognostic indices within a common database has not been routinely performed nor was the reporting of index operating characteristics and major misclassification rates. The RTOG RPA system has been clinically utilized and extensively subjected to validation analyses with no other system demonstrating clear statistical superiority. As the ideal prognostic index has not yet been defined, further research into alternative approaches is warranted.

REFERENCES


