Diagnostic performance of magnetic resonance imaging and computed tomography for advanced retinoblastoma: a systematic review and meta-analysis

Marcus C de Jong, Pim de Graaf, Daniel P Noij, Sophia Göricke, Philippe Maeder, Paolo Galluzzi, Hervé J Brisse, Annette C Moll, Jonas A Castelijns

on behalf of the European Retinoblastoma Imaging Collaboration
ABSTRACT

Purpose • To determine and compare the diagnostic performance of magnetic resonance imaging (MRI) and computed tomography (CT) for the diagnosis of tumor extent in advanced retinoblastoma, using histopathology as reference standard.

Design • Systematic review and meta-analysis.

Participants • Patients with advanced retinoblastoma who underwent MRI and/or CT for the detection of tumor extent from published diagnostic accuracy studies.

Methods • Medline and Embase were searched for literature published until April 2013 assessing the diagnostic performance of MRI and/or CT in detecting intra- and extraorbital tumor extension of retinoblastoma. Diagnostic accuracy data were extracted from included studies. Summary estimates were based on a random-effects model. Intra- and interstudy heterogeneity were analyzed.

Main outcome measures • Sensitivity and specificity of MRI and CT in detecting tumor extent.

Results • Data of the following tumor-extent parameters were extracted: anterior eye segment involvement and ciliary body, optic nerve, choroidal, and (extra-)scleral invasion. Articles on MRI reported results of 591 eyes from 14 studies and CT yielded 257 eyes from 4 studies. The summary estimates with their 95% confidence intervals (CI) of the diagnostic accuracy of conventional MRI at detecting postlaminar optic nerve, choroidal, and scleral invasion showed sensitivities of 59% (95% CI 37–78), 74% (95% CI 52–88), and 88% (95% CI 20–100) respectively and specificities of 94% (95% CI 84–98), 72% (95% CI 31–94), and 99% (95% CI 86–100) respectively. MRI with a high (versus a low) image quality showed higher diagnostic accuracies for detection of prelaminar optic nerve and choroidal invasion, but these differences were statistically not significant. Studies reporting on the diagnostic accuracy of CT did not provide enough data to perform any meta-analyses.

Conclusions • MRI is an important diagnostic tool for the detection of local tumor extent in advanced retinoblastoma although its diagnostic accuracy shows room for improvement, especially sensitivity. With only a few, mostly old studies, there is very little evidence on the diagnostic accuracy of CT and generally these studies show low diagnostic accuracy. Future studies assessing the role of MRI in clinical decision-making in terms of prognostic value for advanced retinoblastoma are needed.
INTRODUCTION

Retinoblastoma is the most frequent malignant ocular tumor in children, typically presenting in the first years of life. It represents approximately 3% of all pediatric malignancies, with an incidence of 1:17,000.\textsuperscript{1}

Retinoblastoma can be accurately diagnosed by fundoscopy and ultrasound, which typically demonstrates an intraocular vascularized and calcified mass. Cross-sectional imaging is primarily used for local tumor staging (related to metastatic risk) and depiction of associated intracranial primitive neuroectodermal tumors (mostly pineoblastoma; see Part II of this thesis).\textsuperscript{2,3} Magnetic resonance imaging (MRI) can also aid retinoblastoma diagnosis in case of an unclear ocular medium. Treatment strategies are focused on survival, preserving vision, and finally on avoiding enucleation. Although the gold standard for diagnosis and local staging relies on pathology, the treatment strategy is frequently based on clinical and radiological findings only, notably in eyes treated conservatively. In patients treated by primary enucleation, extraocular extension must first be ruled out by imaging in order to avoid leaving behind tumor tissue after resection. Therefore, accurate local assessment based on imaging is critical at diagnosis. As recommended nowadays, we perform MRI in all newly diagnosed retinoblastoma patients, but in the past MRI has often been performed as a second diagnostic step and we are not sure about the policies in other institutions.\textsuperscript{4,5}

The most important role of MRI is to aid in the decision to either treat an eye conservatively or to enucleate the eye. The prediction of high-risk features of retinoblastoma based on clinical and radiological features is pivotal to select the best treatment option.\textsuperscript{6} Important risk factors for local recurrence and metastasis are massive choroidal invasion, scleral invasion, optic nerve invasion posterior to the lamina cribrosa (especially if the surgical resection margin of the optic nerve is invaded), and involvement of the anterior eye segment.\textsuperscript{6–12}

Several studies have been published on the diagnostic accuracy of MRI and only a few on computed tomography (CT) for various tumor-extent parameters. CT has been considered important in detecting calcifications, but it has recently been demonstrated that the combination of ultrasound and MRI has the same sensitivity and specificity compared with CT in detecting intratumoral calcifications in retinoblastoma.\textsuperscript{13} Moreover, whenever possible CT should be avoided in young children because it poses a significant radiation risk,\textsuperscript{14–17} and even more so in patients with hereditary retinoblastoma.\textsuperscript{4,13} However, because the incidence of retinoblastoma is low, study populations are generally small. There is also considerable heterogeneity (e.g., in terms of image quality) among the studies. To overcome these problems a critical systematic review of the different studies is required.

The purpose of this study was to provide a complete overview of available evidence and to determine and compare the diagnostic performance of MRI and CT for the detection of tumor extent in advanced retinoblastoma patients, with histopathology as reference standard.
METHODS
We performed this study according to the PRISMA statement for systematic reviews and meta-analyses.18,19

Search strategy
We searched Medline (PubMed) and Embase for English, Dutch, German, and Spanish literature published until April 2013, evaluating tumor extent of retinoblastoma by MRI or CT. We also included alternatively found studies (e.g., through checking references in included studies). When necessary we have contacted authors for additional data. The search included keywords corresponding to the index test MRI and CT, the target condition (retinoblastoma) and diagnostic performance (appendix A).

Study selection
Article titles and abstracts were independently reviewed for eligibility by two authors (MCJ and DPN) and discrepancies were resolved by consensus.

Studies were included if they met all of the following criteria: (1) the study population consisted of retinoblastoma patients, (2) the study assessed diagnostic performance of MRI and/or CT as diagnostic test for tumor invasion into the ocular wall, optic nerve invasion, or anterior eye segment involvement, (3) histopathology was used as the reference standard test, (4) if at least one pair of the absolute numbers of true positives (TP) and false negatives (FN), or true negatives (TN) and false positives (FP) were available or could be derived adequately. For inclusion in the meta-analysis TP, FP, TN, and FN should all four be available.

Studies were excluded if they met one of the following criteria: (1) the article was a review or meta-analysis, (2) (potentially) overlapping study populations were reported for the same outcome.

Diagnostic accuracy of tumor extension into the optic nerve can be assessed in different ways. To avoid unclear definitions, in this study we will discuss three categories: (1) if the optic nerve disk has been invaded by tumor tissue (from here on to be referred to as prelaminar optic nerve invasion), (2) invasion exactly into the lamina cribrosa (from here on to be referred as intralaminar optic nerve invasion), and (3) invasion posterior to the lamina cribrosa (from here on to be referred as postlaminar optic nerve invasion).

Data extraction
Two authors (MCJ and DPN) independently extracted study data. Discrepancies were resolved by consensus. When studies reported multiple sets of sensitivity and specificity from multiple readers separately, the set with the highest diagnostic odds ratio was used for the figures and meta-analysis. We have done this to prevent data from the same patient population to be used twice. This could have resulted in an overestimation of the summary estimates; therefore, we also reported the overall summary estimates of including sensitivity and specificity from the reader with the lowest diagnostic odds ratio.

Risk of bias assessment
Ideally studies only included eyes that were primarily enucleated (i.e., retinoblastoma that was not previously treated), because retinoblastoma treatment can influence the appearance of tumor-extent parameters on magnetic resonance (MR) images. However, some authors have included both secondary and primary enucleations or have not mentioned this at all in
their article. We have performed sensitivity analysis by assessing the effect of analyzing the diagnostic performance of studies that explicitly state having only included eyes that were primary enucleations.

In an empirical study of bias in diagnostic tests Lijmer et al. have demonstrated that study design can be very important to the results of diagnostic tests; they showed that a case-control design caused an overestimation of the diagnostic odds ratio by three times, the use different reference tests overestimated the diagnostic odds ratio two times, and not blinding of the test interpretation overestimated the diagnostic odds ratio by 30%. However, verification bias, a nonconsecutive or random patient selection and a retrospective study design did not seem to affect the diagnostic odds ratio.

We used the QUADAS-2 (quality assessment of studies of diagnostic accuracy included in systematic reviews) checklist to assess the study quality in terms of the risk of bias and the applicability of included studies. Two authors (MCJ and DPN) independently assessed the study quality of the included articles.

Two authors (PdG and HJB), with respectively 11 and 16 years experience in ocular magnetic resonance imaging, independently assessed the image quality (low, intermediate, or high) of MR and CT images provided in the articles, resulting in semi-quantitative quality scores. The guidelines for imaging retinoblastoma by De Graaf et al. served as a checklist for the quality assessment. They recommended imaging of the eyes with a slice thickness of at most 2 mm and a pixel size that is no larger than 0.5×0.5 mm². To achieve this, it is advisable to use either a multi-channel head coil (3.0T system) or surface coils (1.5T system). If information on slice thickness and/or pixel size was not mentioned in the published articles, the included images served as a qualitative measure for image quality. Proper sedation is also important to ensure a high image quality in the usually very young retinoblastoma patient group. Discrepancies between image quality scores were resolved by consensus.

**Data synthesis and statistical analysis**

If tumor-extent parameters were analyzed in enough studies to allow statistical analysis, we analyzed data using a bivariate random effects regression model and provided summary estimates. This model assumes a binomial distribution of the within-study variability (i.e., the variability between sensitivity and specificity within a study). The model assumes correlated normally distributed random effects between studies. The inverse relation between sensitivity and specificity – when the positivity criterion is varied – corresponds to the degree of correlation between the logit sensitivity and logit specificity. We performed metaregression to explore the effect of image quality on the

---

**Figure 1.** Flowchart showing systematic literature search. CT = computed tomography, FN = false-negative results, FP = false-positive results, MRI = magnetic resonance imaging, TN = true-negative results, TP = true-positive results.
diagnostic accuracy of tumor-extent parameters if sample sizes allowed this.

We summarized the data of each study and overall estimates in forest plots with a 95% confidence interval (CI) of sensitivity and specificity for each tumor-extent parameter. We also plotted these numbers in receiver operating characteristic (ROC) spaces showing the summary estimates with a 95% confidence region. Potential outliers have been identified on the basis of the position of individual studies relative to the other studies in these ROC plots. We have performed sensitivity analysis by excluding these potential outliers, but these results should be interpreted with care.

We calculated the diagnostic odds ratio along with sensitivity and specificity as an overall measure of diagnostic performance. The advantage of diagnostic odds ratio is its independence from disease prevalence and approximately normal distribution of the natural logarithm of the diagnostic odds ratio. As a general rule diagnostic tests with a diagnostic odds ratio >25 are considered moderately accurate and tests with a diagnostic odds ratio >100 highly accurate. For the meta-analysis, we used the statistical software packages SAS (Proc NLMIXED, SAS v9.3, Raleigh, NC, USA). We created the forest plots with Photoshop (CS6, San Jose, CA, USA). We used Cochrane's Review Manager (version 5.2, Copenhagen, Denmark) to create the ROC plots.

RESULTS

Medline and Embase searches yielded 426 unique studies. We excluded 376 articles based on title and abstract. We excluded 32 studies based on the full text; see figure 1 for reasons of exclusion. The study characteristics show considerable differences between the included studies (appendix B). Eighteen studies met the inclusion criteria for qualitative synthesis and thirteen studies were included in the meta-analysis. The study population by Schueler et al. overlapped with the more recent study by Lemke et al., therefore we only extracted data from Schueler's study for tumor-extent parameters not included in Lemke's study (i.e., scleral invasion). All analyses were performed on a per eye basis and not a per patient basis, as some studies have included two eyes from one patient (appendix B). Articles on MRI reported results of 591 eyes from 14 studies (excluding Schueler et al.) and articles on CT yielded 257 eyes from 4 studies. The articles reported ages at diagnosis of retinoblastoma patients that ranged from a mean age of 11 to a mean age of 32 months. To avoid very large tables we have split the tumor-extent parameters into three groups: optic nerve invasion, choroidal invasion, and scleral invasion (appendices C–E). Diagnostic accuracy data of MRI at detecting optic nerve invasion, choroidal invasion, and scleral invasion proved to be sufficient for meta-analysis.

The sensitivities and specificities of tumor-extent parameters are presented in appendices C–E. Forest plots show the sensitivities and specificities of each tumor-extent parameter with their 95% CIs as horizontal lines sorted by sensitivity (appendix F).

Risk of bias assessment

See figures 2 and 3 for the QUADAS summary scores regarding risk of bias and concern of applicability of the included studies. The entire list of QUADAS scores for each study is available in the electronic supplement (appendix G).
None of the included studies had a case-control design. All studies had the same reference standard, as histopathology was one of the inclusion criteria of this systematic review. MRI interpretation was blinded in ten studies, but this was unclear in the other eight. Histopathological interpretation was blinded in eight studies, while this was unclear in ten studies. We scored seven out of eighteen studies as having an appropriate time interval between the MRI and enucleation, and the other eleven studies as unclear, because they either did not report a time interval at all or because they reported relatively wide ranges (appendices B and G).

Only the most severe retinoblastoma cases are included, since enucleation is necessary for histopathological assessment, but we don’t know how this affects the applicability of the results on patients with less severe cases.

### Table 1. Summary estimates of diagnostic performance of MRI for retinoblastoma tumor extent

<table>
<thead>
<tr>
<th>TE</th>
<th>Sensitivity*</th>
<th>Specificity*</th>
<th>Diagnostic odds ratio*</th>
<th>No. of studies</th>
<th>Disease prevalence†</th>
<th>Number of true and false test results in a sample of 1000 patients‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prelaminar</td>
<td>Overall</td>
<td>73 (45–90)</td>
<td>96 (64–100)</td>
<td>8</td>
<td>47 (157/334)</td>
<td>342 23 507 128</td>
</tr>
<tr>
<td></td>
<td>Only primary enucleations§</td>
<td>80 (31–79)</td>
<td>81 (3–100)</td>
<td>4</td>
<td>70 (119/169)</td>
<td>564 55 241 140</td>
</tr>
<tr>
<td></td>
<td>Worst reader*</td>
<td>70 (43–88)</td>
<td>95 (62–100)</td>
<td>8</td>
<td>47 (157/334)</td>
<td>329 24 506 141</td>
</tr>
<tr>
<td>Intralaminar</td>
<td>Overall</td>
<td>38 (0–100)</td>
<td>89 (47–99)</td>
<td>17 (36/217)</td>
<td>62 92 742 104</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Only primary enucleations§</td>
<td>16 (0–100)</td>
<td>84 (11–100)</td>
<td>3</td>
<td>19 (35/181)</td>
<td>31 128 679 162</td>
</tr>
<tr>
<td></td>
<td>Worst reader*</td>
<td>39 (1–100)</td>
<td>89 (44–99)</td>
<td>4</td>
<td>17 (36/217)</td>
<td>64 94 740 102</td>
</tr>
<tr>
<td>Postlaminar</td>
<td>Overall</td>
<td>59 (37–78)</td>
<td>94 (84–98)</td>
<td>20.9 (4.75–92.2)</td>
<td>10 17 69 (418)</td>
<td>17 53 781 68</td>
</tr>
<tr>
<td></td>
<td>Only primary enucleations§</td>
<td>56 (27–81)</td>
<td>93 (77–98)</td>
<td>16.0 (2.26–113)</td>
<td>7</td>
<td>18 (59/331) 100 61 761 78</td>
</tr>
<tr>
<td></td>
<td>Worst reader*</td>
<td>58 (31–81)</td>
<td>94 (85–97)</td>
<td>19.9 (4.64–85.4)</td>
<td>10</td>
<td>17 (69/418) 95 53 782 70</td>
</tr>
<tr>
<td>Choroid</td>
<td>Overall</td>
<td>74 (52–88)</td>
<td>72 (31–94)</td>
<td>7.38 (1.21–45.0)</td>
<td>6</td>
<td>42 (105/253) 309 165 420 107</td>
</tr>
<tr>
<td></td>
<td>Only primary enucleations§</td>
<td>69 (26–93)</td>
<td>78 (6–100)</td>
<td>7.73 (0.07–854)</td>
<td>4</td>
<td>44 (81/184) 302 123 437 139</td>
</tr>
<tr>
<td>Sclera</td>
<td>Overall</td>
<td>88 (20–100)</td>
<td>99 (86–100)</td>
<td>503 (24.9–1.02×10⁴)</td>
<td>6</td>
<td>4 (12/281) 38 14 944 5</td>
</tr>
<tr>
<td></td>
<td>Only primary enucleations§</td>
<td>IC</td>
<td>IC</td>
<td>2</td>
<td>5 (6/133)</td>
<td>IC IC IC IC</td>
</tr>
</tbody>
</table>

TE = tumor extent, FN = false negative, FP = false positive, IC = incalculable, TN = true negative, TP = true positive.
*Confidence intervals are in parentheses. Sensitivity, specificity and disease prevalence are percentages.
†The disease prevalence is defined by true positives and false negatives divided by the total sample size ([TP+FN]/[TP+FP+TN+FN] in parentheses).
‡Based on the sensitivity, specificity and disease prevalence of the tumor-extent parameter; sometimes the total sample size is 999 or 1001 due to rounding.
§Studies that reported explicitly that their data was based on primary enucleations only (appendices B and C).
°Data from the worst reader in terms of diagnostic odds ratio in case studies reporting multiple sets of sensitivity and specificity from different readers (appendix C). All other analyses include data from the best reader.

### Figure 2. Graph showing the risk of bias and applicability concerns: review of authors’ judgements about each domain, presented as percentages across included studies.
severe retinoblastoma. Three studies raised applicability concerns for the index test: Lee et al. reported using four different MR scanners, Brisse et al. reported using various different MR and CT scanners and John-Mikolajewski et al. report using contrast enhancement in only three out of eleven patients. Ainbinder et al. didn’t describe the reference test in their study leading to an unclear concern of the applicability (figure 3).

Only nine studies explicitly report that patients were treated with primary enucleations (i.e., without other retinoblastoma treatment before enucleation), one study mentions the use of pre-operative chemotherapy in some patients but does not stratify the results, two studies are unclear about this issue and six studies didn’t report on this at all (appendix B).

Summary estimates

Compared with histopathology the meta-analysis of MRI for the different tumor-extent parameters gave us a sensitivity of 73% (95% CI 45–90), a specificity of 96% (95% CI 64–100), and a diagnostic odds ratio of 59.7 (95% CI 4.36–818) for prelaminar optic nerve invasion from 8 studies (table 1 and appendix F). For intralaminar optic nerve invasion, MRI showed a sensitivity of 38% (95% CI 0–100), a specificity of 89% (95% CI 47–99), and a diagnostic odds ratio of 4.84 (95% CI: 0.00–3.14×10⁴) from 4 studies. For postlaminar optic nerve invasion, MRI demonstrated a sensitivity of 59% (95% CI 37–78), a specificity of 94% (95% CI 84–98), and a diagnostic odds ratio of 20.9 (95% CI 4.75–92.2) from 10 studies. For choroidal invasion, MRI showed a sensitivity of 74% (95% CI 52–88), a specificity of 72% (95% CI 31–94), and a diagnostic odds ratio of 7.38 (95% CI 1.21–45.0) from 6 studies. For scleral invasion MRI demonstrated a sensitivity of 88% (95% CI 20–100), a specificity of 99% (95% CI 86–100), and a diagnostic odds ratio of 503 (95% CI 24.9–1.02×10⁴) from 6 studies. The ROC plots show the summary estimates of specificity on the x-axis and sensitivity on the y-axis for these four tumor-extent parameters with their respective 95% confidence areas (figure 4).

Sensitivity analysis and metaregression

Based on the ROC plots we have identified two studies as potential outliers: the study by Wilson et al. for postlaminar optic nerve invasion and the study from Khurana et al. for choroidal invasion (figure 4). After exclusion of these potential outliers the diagnostic accuracy of MRI for postlaminar optic nerve invasion showed a sensitivity of 61% (95% CI 32–84), a specificity of 96% (95% CI 80–99), and a diagnostic odds ratio of 34.6 (95% CI 5.80–206). For choroidal invasion MRI showed a sensitivity of 77% (95% CI

---

**Figure 3.** Chart summarizing the risk of bias and applicability concerns: review of authors’ judgements about each domain for each included study.
Diagnostic performance of retinoblastoma imaging

We have performed a sensitivity analysis of including data from certain primary enucleations only. For prelaminar optic nerve invasion four out of eight studies, for intralaminar optic nerve invasion three out of four studies, for postlaminar optic nerve invasion seven out of ten studies, for choroidal invasion four out of six studies, and for scleral invasion two out of six studies explicitly reported data based on primary enucleations only. Results based on this subset gave lower diagnostic odds ratios for most tumor-extent parameters, only the diagnostic odds ratio of choroidal invasion did not show much of a difference, and for scleral invasion we could not calculate summary estimates (table 1).

We have also assessed the effect of including data from the readers of the MR images with the lowest diagnostic odds ratio – from studies that report multiple readers – on the summary estimates. Three out of eight studies assessing prelaminar optic nerve invasion, one out of four assessing intralaminar optic nerve invasion, and two out of ten assessing postlaminar optic nerve invasion reported two sets of sensitivity and specificity from two readers. When data from the worst reader (i.e., with the lowest diagnostic odds ratio) were included in the analysis the results did not change much (table 1).

For prelaminar optic nerve invasion and choroidal invasion the results from metaregression analysis showed higher diagnostic odds ratios for studies with a higher image quality. The results for postlaminar optic nerve invasion only showed a higher diagnostic odds ratio after removal of a potential outlier. However, none of these differences were statistically significant (table 2 and appendix F).

Table 2. Metaregression of the effect of image quality on diagnostic performance of MRI

<table>
<thead>
<tr>
<th>TE</th>
<th>Sensitivity*</th>
<th>Specificity*</th>
<th>Diagnostic odds ratio*</th>
<th>No. of studies</th>
<th>Disease prevalence†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prelaminar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low quality</td>
<td>57 (27–83)</td>
<td>97 (55–100)</td>
<td>47.0 (1.51–1.47×10³)</td>
<td>4</td>
<td>43 (58/135)</td>
</tr>
<tr>
<td>High quality</td>
<td>83 (58–95)</td>
<td>93 (38–100)</td>
<td>67.2 (2.74–1.65×10³)</td>
<td>4</td>
<td>50 (99/199)</td>
</tr>
<tr>
<td>P values</td>
<td>0.12</td>
<td>0.59</td>
<td>0.85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postlaminar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low quality</td>
<td>59 (32–82)</td>
<td>94 (78–98)</td>
<td>21.1 (3.01–148)</td>
<td>5</td>
<td>21 (44/210)</td>
</tr>
<tr>
<td>High quality</td>
<td>58 (27–84)</td>
<td>94 (77–98)</td>
<td>20.5 (2.41–175)</td>
<td>5</td>
<td>12 (25/208)</td>
</tr>
<tr>
<td>P values</td>
<td>0.94</td>
<td>0.98</td>
<td>0.98</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choroid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low quality</td>
<td>71 (7–99)</td>
<td>66 (16–95)</td>
<td>4.84 (0.10–246)</td>
<td>3</td>
<td>42 (64/151)</td>
</tr>
<tr>
<td>High quality</td>
<td>77 (52–92)</td>
<td>76 (21–97)</td>
<td>10.9 (0.93–128)</td>
<td>3</td>
<td>40 (41/102)</td>
</tr>
<tr>
<td>P values</td>
<td>0.80</td>
<td>0.72</td>
<td>0.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Choroid</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low quality</td>
<td>IC</td>
<td>IC</td>
<td>IC</td>
<td>2</td>
<td>36 (24/66)</td>
</tr>
</tbody>
</table>

TE = tumor extent, FN = false negative, FP = false positive, IC = incalculable, TN = true negative, TP = true positive. P values are for low quality versus high quality. Sensitivity, specificity, and disease prevalence are percentages. *Confidence intervals are in parentheses. Sensitivity, specificity and disease prevalence are percentages. †The disease prevalence is defined by true-positive results and false-negative results divided by the total sample size ([TP+FN]/[TP+FP+TN+FN] in parentheses). §Includes studies with a low image quality score. °Includes studies with either an intermediate or a high image quality score (appendix B). ∧Without data from Wilson et al. 29 #Without data from Khurana et al. 30

CI 41–94), a specificity of 78% (95% CI 46–94), and a diagnostic odds ratio of 11.8 (95% CI 2.14–65.4).
We were able to perform a meta-analysis of diagnostic accuracy of MRI for detection of optic nerve, choroidal, and scleral invasion in retinoblastoma. Study data of extra-scleral invasion, anterior eye segment involvement, and ciliary body invasion proved insufficient for meta-analysis. Studies reporting on the diagnostic accuracy of CT did not provide enough data to perform any meta-analyses.

**Optic nerve invasion**

Postlaminar optic nerve invasion is considered to be one of the most important risk factor for metastasis and has been studied most extensively of all tumor-extent parameters included in this review. The summary estimates of the diagnostic accuracy of MRI are far from perfect, especially in terms of sensitivity (table 1 and appendix F). Increasing image quality seems to have a positive effect on the diagnostic accuracy of MRI, especially in terms of sensitivity (table 2).

Prelaminar optic nerve invasion is not considered to be a risk factor, but we believe the higher diagnostic accuracy of MRI techniques with a higher image quality does show the potential of high-resolution MRI (table 2 and appendix F).

Four studies have also looked at tumor extension exactly into the lamina cribrosa. Because
intralaminar tumor invasion can develop into postlaminar tumor invasion in the future or because a false-negative result could actually be histopathologically a postlaminar invasion, this is a category that deserves attention. MRI shows a reasonable specificity in detecting intralaminar optic nerve invasion, but the reported sensitivities show a wide range (table 1 and appendix F).

High-resolution imaging techniques are critical for borderline postlaminar invasion. The relatively high number of false negatives could be caused by missing minimal postlaminar invasions. We have extracted the degree of tumor involvement per study (unfortunately this was scarcely reported) for TP and FN in millimeters posterior to the lamina cribrosa to illustrate this problem (appendix C). False positives, on the other hand, can be caused by posterior bulging of the lamina cribrosa secondary to an increased intra-ocular pressure.

Other causes of false positives might be caused by inflammation or endothelial proliferation, which can mimic (residual) tumor invasion of the optic nerve.

The sensitivities (0-43%) of postlaminar optic nerve invasion on CT are quite low, even in patients with an extensive involvement (appendix C). Jaquemin et al. showed that non-visualisation of central retinal vessels resulted in a sensitivity of 100% and a specificity of 73%. However, Brisse et al. showed no significant correlation between non-visualisation of central retinal vessels and optic nerve involvement (P=0.65).

**Choroidal and (extra-)scleral invasion**

Like with postlaminar optic nerve invasion high-resolution imaging is mandatory to assess subtle invasions into the choroid. Summary estimates of MRI in detecting choroidal invasion also indicate room for improvement; studies with a higher image quality show a higher diagnostic accuracy (tables 1 and 2 and appendix F). Massive choroidal invasion is considered to be a risk factor for metastasis. However none of the included studies reported diagnostic accuracy for massive choroidal invasion separately. Only Chawla et al. mentioned that 19 of 45 patients with choroidal invasion had massive choroidal invasion, defined as at least 3 mm of tumor ingrowth or reaching scleral tissue by the International Retinoblastoma Staging Work Group. As expected (extra-)scleral invasion is easier to detect or exclude on MRI, showing a higher diagnostic accuracy (table 1 and appendices D and F).

Evidence for diagnostic accuracy of CT is scarce and incomplete. Olivecrona et al. report that choroidal invasion is not visible on CT images, but they did not mention the total number of positive choroidal invasions on histopathology. In the same study they reported a sensitivity of 71% and a specificity of 98% for scleral invasion, but 8 of 10 true-positive scleral invasions extended profoundly into the orbit (appendices D and F).
**Chapter 1**

**Anterior eye segment involvement and ciliary body invasion**

Contrast enhancement of the anterior eye segment on MR images gives sensitivities of 68–93% and specificities of 43–82% for the diagnosis of iris neoangiogenesis in three studies (appendices E and F). De Graaf et al.\(^3\) showed a sensitivity of 100% and a specificity of 88% for MRI compared to anterior eye segment abnormalities visible with ophthalmoscopy. De Graaf et al.\(^4\) also found that the degree of anterior eye segment enhancement on MRI matched angiogenesis in the iris immunohistochemically (P=0.09). Anterior eye segment enhancement has also been found to correlate with optic nerve invasion.\(^{32,41,42}\)

Diagnosis of ciliary body involvement with MRI compared to histopathology shows sensitivities of 71–100% and specificities of 65–100% in a limited number of studies (appendices E and F).

**Limitations**

Because the incidence of retinoblastoma is low, published studies usually have small patient populations, resulting in summary estimates with wide confidence intervals.

For ideal diagnostic test comparison retinoblastoma patients with comparable severity of disease undergo the gold standard test (histopathology) regardless of the index test (MRI of CT). When the index test influences the choice whether the gold standard test will be performed, this is called verification bias. However, a study design where this does not happen is unattainable in practice due to ethical considerations. In this case enucleation depends on previous clinical tests, but also by the index test. Verification bias can lead to overestimation of sensitivity and underestimation of specificity, but generally does not influence the diagnostic odds ratio.\(^{20,43}\) Verification bias is likely to be present in all included studies with varying degree. Studies report varying prevalences of tumor-extent parameters (appendices C–E). Prevalence depends on more than verification bias alone, but a higher than normal prevalence might partly be attributed to verification bias. Inherently the results presented in this meta-analysis are based on patients with severe retinoblastoma, as only these patients undergo enucleation allowing for a histopathological evaluation. To what extent these results are also generalizable to the entire retinoblastoma patient population is difficult to say.

Studies have not always explicitly reported that the interpretation of the MRI or histopathology was performed in a blinded manner. Not blinding the interpretation can result in a slight overestimation of the diagnostic odds ratio.\(^{20}\) Most studies were unclear on the fact if their patients were enrolled consecutively or randomly (yes: 4 studies, unclear: 11 studies, no 1 study) and most studies either reported a retrospective study design or did not mention this in their article (retrospective: 9 studies, not specified: 7 studies, prospective: 2 studies), however, Lijmer et al.\(^{20}\) have shown that these study design aspects don’t impact the diagnostic odds ratio of a diagnostic test. Some studies included both affected eyes of one patient, which might have caused an overestimation of diagnostic accuracy; this issue was not part of the QUADAS checklist, see appendix B for studies that included more eyes than patients.

In meta-analyses there is a risk that publication bias affects the results. Publication bias usually occurs when small negative studies (i.e., studies with poor diagnostic accuracy and a low number of patients) are deemed uninteresting and are not published. Missing these studies could lead to an overestimation of the summary estimates. We have chosen not to look at publication bias formally, because of the lack of a good test.\(^{44}\)
Heterogeneity across studies is a common limitation of meta-analyses. The studies we included differ in terms of imaging technique and quality, prevalence of tumor-extent parameters, inclusion criteria (some also included secondary enucleations), and time between index and reference test. Sensitivity analysis by excluding studies that did not explicitly report only including primary enucleations showed slightly lower diagnostic accuracy for most tumor-extent parameters, but the analysis was done on a smaller subset of studies resulting in even wider confidence intervals (table 1). We were able to analyze the effect of image quality to some extent. We used a random effects model, which adjusts the summary estimates and confidence intervals for between-study variations to a certain extent.

Since the consequences of an incorrect positive or negative test are large, it is important to achieve a higher diagnostic accuracy. Unfortunately not all included studies reported the pixel size and section thickness of their images; therefore we had to use a semi-quantitative quality score to distinguish between low and high-quality studies. Most studies reported the use of sedation or anesthesia, but some studies did not (appendix B). Even in patients under general anesthesia motion artifacts can pose a problem and decrease the image quality. Even though the evidence is not very strong, our results (for choroidal and prelaminar tumor invasion and for postlaminar invasion after exclusion of one potential outlier) did show a positive impact of image quality on diagnostic accuracy. These results were not statistically significant. Nevertheless we believe that high-quality images have the potential to increase diagnostic accuracy considerably and therefore, we think it is important that MRI protocols are at least on par with the most recently published guidelines by De Graaf et al.4, especially in patients with (advanced) retinoblastoma for whom conservative management is considered.

Conclusion
This systematic review gives an extensive overview of the available evidence of the diagnostic accuracy of tumor extent in advanced retinoblastoma. With only a few, mostly old studies, there is very little evidence on the diagnostic accuracy of CT and generally these studies show low diagnostic accuracy. For MRI there is room for improvement, especially for the test sensitivity. Improvement of the diagnostic accuracy of MRI for retinoblastoma tumor-extent staging is pivotal for personalizing retinoblastoma treatment, e.g., for the justification of conservative eye-sparing treatment strategies and for the determination of correct surgical margins in case of unavoidable enucleation. Diagnostic accuracy studies of retinoblastoma tumor extent inherently suffer from a great deal of verification bias overestimating sensitivity and underestimating specificity. Clinical decisions are, however, not just based on the test outcome of one tumor-extent parameter, but rather on a combination of radiological and clinical factors. How these results reflect the value of MRI in the entire clinical decision-making process and to what extent more accurate imaging improves retinoblastoma care remains difficult to predict. Larger prospective studies addressing the role of cross-sectional imaging in the entire clinical decision-making process – including its effect on patient and eye survival – are needed in the future. These results could in turn be used for the development of clear decision-making criteria.
REFERENCES


