Radiological staging in patients with hilar cholangiocarcinoma: a systematic review and meta-analysis

[1]A T RUYS, MD, [2]B E VAN BEEM, MD, [2]M R W ENGELBRECHT, MD, PhD, [2]S BIPAT, PhD, [2]J STOKER, MD, PhD and [1]T M VAN GULIK, MD, PhD

[1]*Department of Surgery, Academic Medical Center, Amsterdam, Netherlands, and* [2]*Department of Radiology, Academic Medical Center, Amsterdam, Netherlands*

Objective: To obtain diagnostic performance values of CT, MRI, ultrasound and 18-fludeoxyglucose positron emission tomography (PET)/CT for staging of hilar cholangiocarcinoma.

Methods: A comprehensive systematic search was performed for articles published up to March 2011 that fulfilled the inclusion criteria. Study quality was assessed with the quality assessment of diagnostic accuracy studies tool.

Results: 16 articles (448 patients) were included that evaluated CT (*n*=11), MRI (*n*=3), ultrasound (*n*=3), or PET/CT (*n*=1). Overall, their quality was moderate. The accuracy estimates for evaluation of CT for ductal extent of the tumour was 86%. The sensitivity and specificity estimates of CT were 89% and 92% for evaluation of portal vein involvement, 83% and 93% for hepatic artery involvement, and 61% and 88% for lymph node involvement, respectively. Data were too limited for adequate comparisons of the different techniques.

Conclusion: Diagnostic accuracy studies of CT, MRI, ultrasound or PET/CT for staging of hilar cholangiocarcinoma are sparse and have moderate methodological quality. Data primarily concern CT, which has an acceptable accuracy for assessment of ductal extent, portal vein and hepatic artery involvement, but low sensitivity for nodal status.

Different imaging investigations are currently used for staging of hilar cholangiocarcinoma patients, yet most evidence is available from CT. Although the quality of imaging has improved in recent years, a substantial proportion of tumours are still found to be unresectable during laparotomy, despite extensive pre-operative work [1]. Only about half of the tumours who are surgically explored are ultimately resectable [2]. Hence, correct staging of hilar cholangiocarcinoma remains a challenge.

Criteria for unresectable disease include locally advanced tumour, distant metastases and lymph node metastases beyond the hepatoduodenal ligament. Locally advanced disease is based on extent of infiltration proximally in the biliary ductal tree, the portal vein and the hepatic artery or its branches. Therefore an accurate assessment of resectability of patients with hilar cholangiocarcinoma includes correct evaluation of local, vascular and nodal status, as well as assessment of distant metastases. Accurate staging of vascular status and proximal ductal extent of hilar cholangiocarcinoma is challenging owing to the anatomic complexity of the hilar region, in conjunction with the small tumour size in most cases.

MRI, CT, ultrasound and 18-fludeoxyglucose (FDG) positron emission tomography (PET)/CT are currently used to assess resectability of hilar cholangiocarcinoma. The aim of this study was to systematically review the literature on these four imaging investigations for assessment of ductal extent of the tumour, portal vein involvement, hepatic artery involvement, lymph node status and presence of metastases in patients with hilar cholangiocarcinoma.

Methods and materials

Literature search

A comprehensive search on studies in human subjects was performed by one observer to identify those studies dealing with the diagnostic performance of CT, ultrasound, MRI and FDG-PET/CT for staging of hilar cholangiocarcinoma. The databases used for the literature search included MEDLINE (January 1966 to March 2011), Embase (January 1980 to March 2011) and the Cochrane Database of Systematic Reviews. Details of the search strategy are shown in Table 1. All review articles, letters, comments and case reports (<10 patients) were eliminated. Articles found to be eligible on the basis of their title and abstract were subsequently selected for full manuscript review. We augmented our literature search by manually reviewing the reference lists of identified studies.

Inclusion and exclusion criteria

Two observers independently reviewed all eligible articles for the following inclusion criteria: the study

Address correspondence to: Mr Anthony Theodore Ruys, IW01, Academic Medical Center, Meiwergdreef 9, 1105AZ Amsterdam, Netherlands. E-mail: a.t.ruys@amc.uva.nl

Table 1. Keywords used in the literature search

Category	MEDLINE	Embase
Patients	''cholangiocarcinoma'' OR ''bile duct cancer'' OR ''klatskin'' OR ''Cholangiocarcinoma'' [Mesh] OR ''Bile Duct Neoplasms''[Mesh]	Klatskin tumour OR bile duct carcinoma OR bile duct tumour
Intervention	''computed tomography'' OR ''Tomography, X-Ray Computed''[Mesh] OR ''Magnetic Resonance Angiography''[Mesh] OR ''Cholangiopancreatography, Magnetic Resonance''[Mesh] OR ''Magnetic Resonance Imaging''[Mesh] OR ''Magnetic Resonance Imaging'' OR ''Ultrasonography''[Mesh] OR ''ultrasonography'' OR ''Positron Emission Tomography'' AND (''Neoplasm Staging'' [Mesh] OR ''staging'' OR ''resectability'')	(Computer-assisted tomography OR nuclear magnetic resonance imaging OR positron emission tomography OR echography) AND (staging OR Cancer Staging)

population (hilar cholangiocarcinoma patients with reference standard) included at least 10 patients; presented data were about primary staging; the reference standard included surgical or pathological confirmation; and 2×2 contingency tables or data allowing their reconstruction were presented for portal vein involvement, hepatic artery involvement, lymph node involvement or presence of distant metastasis. Articles giving sufficient data to calculate the accuracy for ductal extent of the tumour were also included, because these data are not presented in 2×2 tables. Details are described in the following section.

Studies were excluded when no specific data of hilar cholangiocarcinoma patients could be retrieved (when patients were pooled with patients with other bile duct carcinomas). When duplicate studies presenting data about the same population were found, only the most recent one was included. If an article presented data on more than one staging element, inclusion and exclusion criteria were checked for each discrete data set. Disagreements on inclusion suitability were resolved by consensus.

Data extraction

Both observers independently extracted relevant data from each article by using a standardised form. The readers were not blinded to the authors, journal or year of publication. Disagreements were resolved by consensus.

Quality assessment

Methodological quality of included studies was assessed independently by the two readers using the quality assessment of diagnostic accuracy studies (QUADAS) tool, which is a quality assessment tool specifically developed for systematic reviews of diagnostic accuracy studies [3]. Nine items were extracted; six items were deemed irrelevant, given the inclusion and exclusion criteria (*e.g.* whether the reference standard was likely to correctly classify the target condition). In addition to the nine QUADAS items, the design (retrospective or prospective) of the study was scored. Criteria for different QUADAS items are shown in Table 2.

Other study characteristics and imaging features

The following data were recorded for each article: sample size, year of publication, male:female ratio, age,

stage of disease, imaging modality (CT, ultrasound, MRI, PET/CT) and technique.

For studies in which CT was used, the following imaging features were recorded: system type, contrast material use, amount of iodine administered, imaging phases, section thickness and number of detection rows. For MRI, we recorded the magnetic field strength, contrast material use, pulse sequences and section thickness. For PET/CT, we recorded the system type, tracer specifics, scanning time and type of analysis. For ultrasound, we recorded the system type, megahertz and experience of the ultrasonographer.

Data for calculating diagnostic values

For evaluation of the portal vein, hepatic artery and lymph node involvement, and distant metastasis, we constructed a 2×2 contingency table on a per-patient basis for the imaging modality used as compared with the reference standard. For ductal extent of the tumour we calculated the accuracy for each study as follows: correct diagnosis of the ductal extent of the tumour (as classified by the Bismuth–Corlette classification [4]) divided by the total number of patients.

Statistical analysis

Accuracy

We calculated the accuracy for each study by dividing the correct staging by the number of patients. We used random-effects models to obtain summary estimates of accuracy. All analyses were performed on logit-transformed accuracy because these are assumed to follow a normal distribution across studies, and therefore the mean logit accuracy corresponding standard errors can be calculated. After antilogit transformation, summary estimates of accuracy with their 95% confidence intervals (CIs) were obtained.

Sensitivity and specificity

For each study, we calculated the sensitivity and specificity on a per-patient basis. We calculated sensitivity as true-positive findings divided by (false-negative findings + true-positive findings) and specificity as true-negative findings divided by (false-positive findings + true-negative findings) on a per-patient basis. We used random bivariate models to summarise estimates of

Table 2. Criteria for quality assessment of diagnostic accuracy studies (QUADAS) items

QUADAS items	Positive score
1. Patient spectrum	Consecutive series with age, sex and stage are mentioned in the publication; without a clear homogeneity in sex, Bismuth–Corlette classification or aetiology (*e.g.* only Bismuth type IV tumours, or primary sclerosing cholangitis patients)
2. Selection criteria	Mentioned in publication
3. Time period between index test and reference standard	Less than 31 days
4. Verification with reference standard	Reference test was performed in all patients or the majority (90%) of the sample
5. Details index test	Sufficient description of the index test to reproduce results, specific variables needed to be described for each modality (*e.g.* imager specifics, section thickness and contrast material use)
6. Details reference standard	Clear description of how surgical and pathological confirmation was acquired.
7. Index results without knowledge	When described that index test results were interpreted by radiologists blinded for the reference standard
8. Reference standard without knowledge	The surgeons were probably never blinded to the index text; hence, ''yes'' when confirmation was based on blinded pathological results
9. Clinical data available	When authors mentioned the use of clinical data; when no references of clinical data were found, we deemed this as unclear
10. Prospective design	Mentioned in publication

sensitivity and specificity with 95% CIs. All analyses were performed on logit-transformed sensitivity and specificity, because these are assumed to follow a normal distribution across studies, and therefore the mean logit sensitivity and specificity with corresponding standard errors were obtained. After antilogit transformation, we obtained summary estimates of sensitivity and specificity along with their 95% CIs. These analyses were executed by using procNlmixed in SAS® v. 9.2 software (SAS Institute, Cary, NC).

Results

Study selection

The initial search yielded 766 articles (Figure 1). The Cochrane Database of Systematic Reviews yielded no additional studies. A total of 135 were duplicate studies, and were excluded. An additional 576 articles were excluded based on review of article titles, abstracts or both. The most frequent reason for exclusion was that the study concerned a review, comment or case report. We excluded another four studies, since these were written in Chinese and we could not read the manuscript.

The manual search of cross-references of the 51 remaining studies yielded an additional 19 articles based on title, and subsequently the abstract. Of the 70 full-text publications, 54 were excluded. 20 publications presented the data in a format that precluded construction of 2×2 tables, and 22 publications included a wider spectrum of bile duct cancers, and specific data about hilar cholangiocarcinoma could not be retrieved. Other reasons for exclusion were: data not about primary staging ($n=3$), <10 hilar cholangiocarcinoma patients ($n=1$), no surgical or pathological reference standard ($n=6$) and duplicate studies ($n=2$). The remaining 16 articles met the inclusion criteria.

Quality assessment

Table 3 shows methodological assessment of included studies using the QUADAS checklist. There was a large variation in methodological quality, and many QUADAS items were not adequately described in the publication. Hence, many QUADAS items were scored as unclear. A consecutive series of patients with a description of age, sex and classification was presented in six studies. Inclusion and exclusion criteria were clearly mentioned in seven studies. The time period between imaging and the reference standard was mentioned in only 9 studies, and was 31 days or less in 7 studies. In none of the studies was the reference standard established by blinded pathological assessment, and therefore no reference standard was obtained without knowledge of the index test. In most studies, it was unclear whether the same clinical information typically available in clinical practice was available at the time of the index test interpretation. Only one study reported a prospective design. Eight studies were retrospectively performed, and in seven studies the design was unclear.

Study characteristics

The characteristics of the included studies are presented in Table 4. The total number of patients in a study ranged from 11 to 83 (median 22 patients). Reported age ranged from 29 to 89 years, and the proportion of male patients ranged from 46% to 91%. Most studies evaluated the performance of CT ($n=11$); other studies evaluated the use of MRI ($n=3$), ultrasound ($n=3$) and PET/CT ($n=1$). Two small studies compared two investigations (CT and MRI) head to head [5, 6]. The included studies evaluated only some of the staging items (ductal extent, portal vein, hepatic artery, nodal status and metastasis), as is shown in Table 5.
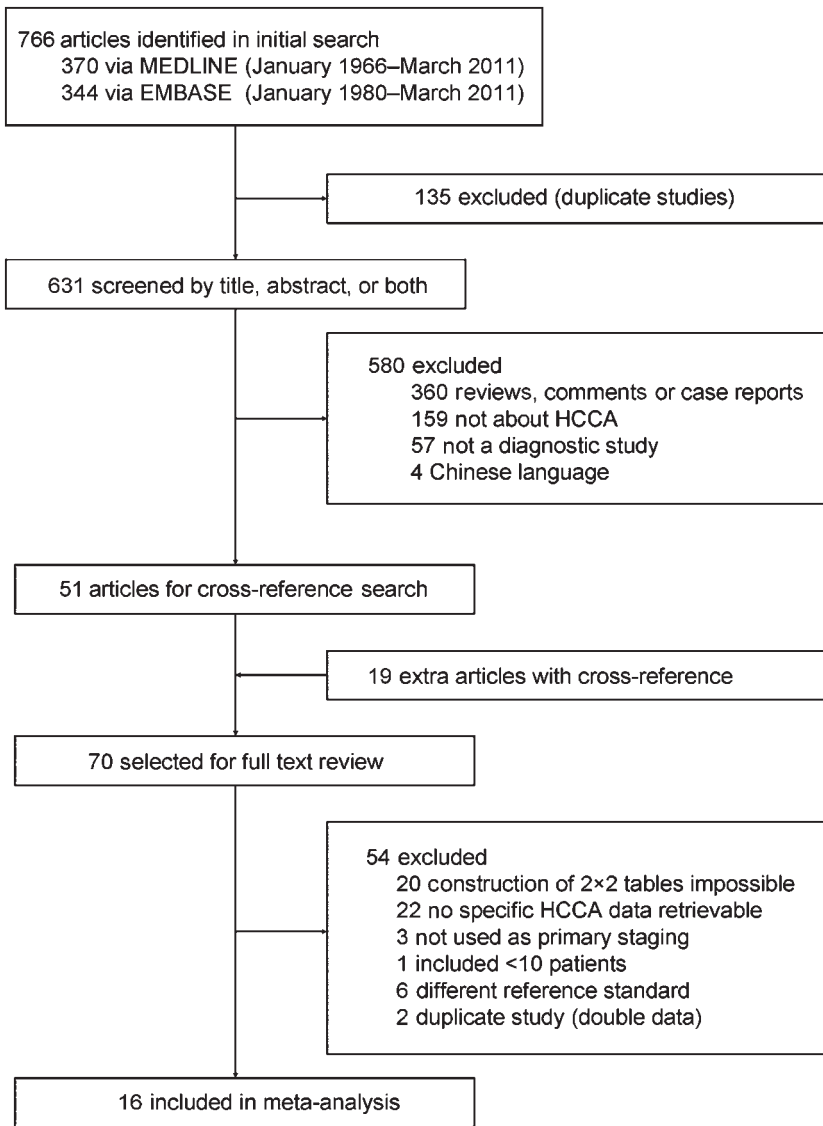
Figure 1. Flow diagram of articles included in the meta-analysis. HCCA, hilar cholangiocarcinoma.

Table 3. Evaluation of included studies using the quality assessment of diagnostic accuracy studies (QUADAS) tool

Source	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Looser et al 1992 [7]	U	−	+	U	U	−	U	−	U	−
Kim et al 2010 [6]	U	+	+	−	−	+	+	−	−	U
Neumaier et al 1995 [10]	U	U	U	−	−	U	U	−	U	U
Chen et al 2009 [18]	+	+	+	+	+	U	U	−	+	+
Akamatsu et al 2010 [12]	+	U	U	+	+	+	+	−	U	−
Watadani et al 2008 [19]	U	+	U	−	−	−	+	−	−	−
Li et al 2008 [11]	U	U	U	+	+	+	U	−	U	U
Cho et al 2007 [5]	U	+	U	−	+	U	+	−	−	−
Endo et al 2007 [20]	+	+	U	−	+	+	U	−	U	U
Feydy et al 1999 [21]	+	+	+	−	+	+	+	U	+	U
Hann et al 1997 [8]	+	U	+	U	+	U	U	−	U	−
Lee et al 2006 [22]	U	+	−	+	+	+	+	−	+	−
Unno et al 2007 [23]	+	U	U	U	+	+	+	U	U	−
Masselli et al 2008 [9]	U	+	+	−	+	+	+	−	+	−
Sugiura et al 2008 [24]	U	U	+	U	+	+	+	−	−	U
Engels et al 1989 [25]	−	U	U	U	−	U	U	−	U	U

+, yes; −, no; U, unclear.
Q1–10=QUADAS items; see Table 2.

Imaging features and evaluation

CT was performed in 11 studies (no system type was specified in one study, a single-slice helical scanner was used in another study, the remaining 9 studies used a multislice helical scanner). Iodinated contrast material was intravenously administered in all studies, with the possible exception of one study that did not report any technical details of the CT. 8 of the 11 studies used a 16-slice CT. The type of contrast material was reported in four studies, with variation in the administered amount (100–150 ml). All studies used three or four phases, including a portal and an arterial phase. The delay after injection of contrast agent before image acquisition varied in the ranges 5–25 s for the arterial phase and 20–70 s for the portal phase. The section thickness was described in eight studies (range 1.25–6.00 mm).

MRI was performed in three studies, including one that did not report details on the MRI. The other two studies used 1.5 T MRI equipment. One study used intravenous gadolinium contrast enhancement; no specific contrast media were used.

PET/CT was evaluated in one study that used a combination of a single-slice spiral CT and full-ring PET. PET imaging data were acquired 60 min after injection of 350 MBq of FDG. PET images were corrected for attenuation based on the CT data.

Ultrasound was performed in three studies using 2.5–5.0 MHz systems. Only one study reported the experience of the sonographers performing the ultrasound, which was 5 and 20 years in this study.

Table 5. Included studies about staging of hilar cholangiocarcinoma

Parameter	CT	MRI	Ultrasound	PET/CT
Ductal extent	8 (238)	3 (49)	2 (61)	—
Portal vein	7 (274)	1 (15)	2 (45)	—
Hepatic artery	6 (191)	—	2 (46)	—
Lymph node	5 (136)	—	—	1 (17)
Metastasis	1 (22)	—	—	1 (17)

PET, positron emission tomography.
Data are the number of studies, with the total number of patients in parentheses.

Diagnostic performance

Ductal involvement

Most CT studies provided data about the assessment of ductal extent of the tumour. Table 6 shows accuracy for ductal involvement of the various studies. The summary estimate for accuracy for CT was 86% (95% CI, 77–92%).

Three included studies reported on the performance of MRI. Accuracy for ductal extent for MRI varied from 71% to 80%.

Two studies were found reporting on the performance of ultrasound for the assessment of ductal extent of the tumour, one with an accuracy of 59% (13/22) [7] and the other with an accuracy of 82% (32/39) [8].

No data on PET/CT were available.

Table 4. Characteristics of included studies

Source	Number of patients	Number of male patients (%)	Mean age (range), years	Modality	Design	Details	Maximum time imaging to surgery (days)
Looser et al 1992 [7]	22	13 (59)	61 (41–75)	Ultrasound	Retro	3.5 and 5 MHz	10
Hann et al 1997 [8]	39	18 (46)	64 (47–82)	Ultrasound	Retro	2.5–5 MHz, 5 or 20 year experience	22
Neumaier et al 1995 [10]	23	25 (71)[a]	62 (41–89)	Ultrasound	Retro	3.5 MHz	NA
Chen et al 2009 [18]	75	68 (91)	61	CT	Pros	16-slice MDCT, 3 phases	14
Akamatsu et al 2010 [12]	22	10 (46)	71 (51–88)	CT	Retro	16-slice MDCT, 3 phases	NA
Watadani et al 2008 [19]	13	NA	NA	CT	Retro	4-, 8- and 16-slice MDCT, 3 phases	NA
Endo et al 2007 [20]	15	13 (65)[a]	65 (50–80)	CT	Unclear	16-slice MDCT, 4 phases	NA
Feydy et al 1999 [21]	11	8 (72)	52 (41–64)	CT	Retro	Helical CT	21
Lee et al 2006 [22]	55	37 (67)	59 (29–76)	CT	Retro	Single and MDCT, 3 phases	42
Unno et al 2007 [23]	24	21 (88)	64	CT	Retro	16-slice MDCT, 4 phases	NA
Sugiura et al 2008 [24]	83	57 (69)	63 (30–80)	CT	Unclear	16-slice MDCT, 4 phases	30
Engels et al 1989 [25]	22	NA	NA	CT	Unclear	NA	NA
Masselli et al 2008 [9]	15	11 (73)	58 (49–74)	MRI	Retro	1.5 T, gadolinium contrast	10
Cho et al 2007 [5]	14	21 (64)[a]	61 (37–84)	CT, MRI	Retro	16-slice MDCT, 3 phases, 1.5 T, RARE and HASTE	NA
Kim et al 2010 [6]	20	11 (55)	64	CT, MRI	Retro	16- or 64-slice MDCT, 3 phases	30
Li et al 2008 [11]	17	11 (65)	62	PET/CT	Unclear	Single-slice spiral CT, full-ring FDG-PET corrected for attenuation	NA

FDG, 18-fludeoxyglucose; MDCT, multidetector CT; NA, not available; PET, positron emission tomography; pros, prospective; retro, retrospective.
[a]Not all patients were included in the analysis of staging elements.

Table 6. Diagnostic accuracy of CT, MRI and ultrasound for assessment of ductal extent of the tumour

Source	Ductal extent correct/total patients (accuracy, %)
CT	
Akamatsu et al 2010 [12]	21/22 (95)
Kim et al 2010 [6]	16/20 (80)
Chen et al 2009 [18]	72/75 (96)
Watadani et al 2008 [19]	12/13 (92)
Endo et al 2007 [20]	13/15 (87)
Lee et al 2006 [22]	46/55 (84)
Unno et al 2007 [23]	17/24 (71)
Cho et al 2007 [5]	9/14 (64)
Summary estimates (95% CI)	86 (77–92)
MRI	
Masselli et al 2008 [9]	12/15 (80)
Kim et al 2010 [6]	15/20 (75)
Cho et al 2007 [5]	10/14 (71)
Ultrasound	
Looser et al 1992 [7]	13/22 (59)
Hann et al 1997 [8]	32/39 (82)

CI, confidance interval.

Portal vein involvement

7 studies, including 292 patients, reported sensitivity and specificity as shown in Table 7. Summary estimates were 89% for sensitivity (95% CI, 80–94%) and 92% for specificity (95% CI, 85–96%).

Only one study evaluated the use of MRI for assessment of portal vein involvement and found a sensitivity of 79% (11/14) and specificity of 0% (0/1) [9].

Two studies evaluated the use of ultrasound for assessing portal vein involvement, and reported a sensitivity and specificity of 75% (6/8) and 93% (12/13) [7], and 83% (15/18) and 100% (5/5) [10], respectively.

No data on PET/CT were available.

Hepatic artery involvement

6 studies, including 191 patients, reported sensitivity and specificity on hepatic artery involvement evaluated by CT (Table 7). Summary estimates were 84% for sensitivity (95% CI, 63–94%) and 93% for specificity (95% CI, 69–99%).

Two studies reported on the use of ultrasound for assessment of involvement of the hepatic artery, and reported a sensitivity and specificity of 0% (0/1) and 100% (21/21) [7], and 43% (3/7) and 100% (4/4) [10], respectively.

No studies were identified that evaluated the use of MRI or PET/CT for assessing hepatic artery involvement.

Table 7. Diagnostic accuracy of CT for portal vein, hepatic artery and lymph node involvement

Source	Modality	True-positive	False-positive	False-negative	True-negative	Sensitivity, percentage (95% CI)	Specificity, percentage (95% CI)
Portal vein							
Chen et al 2009 [18]	CT	39	0	3	33	93 (81–93)	100 (89–100)
Sugiura et al 2008 [24]	CT	20	7	1	55	95 (76–100)	89 (78–95)
Lee et al 2006 [22]	CT	20	2	6	27	77 (56–91)	93 (77–99)
Akamatsu et al 2010 [12]	CT	7	2	1	12	88 (47–100)	86 (57–98)
Endo et al 2007 [20]	CT	5	2	0	8	100 (48–100)	80 (44–98)
Feydy et al 1999 [21]	CT	5	0	2	4	71 (29–96)	100 (40–100)
Watadani et al 2008 [19]	CT	4	0	0	9	100 (40–100)	100 (66–100)
Summary estimates	CT					89 (80–94)	92 (85–96)
Masselli et al 2008 [9]	MRI	11	1	3	0	79 (49–95)	0 (0–98)
Looser et al 1992 [7]	Ultrasound	6	1	2	13	75 (35–97)	93 (66–100)
Neumaier et al 1995 [10]	Ultrasound	15	0	3	5	83 (59–96)	100 (48–100)
Hepatic artery							
Chen et al 2009 [18]	CT	25	0	5	45	83 (65–94)	100 (92–100)
Lee et al 2006 [22]	CT	19	1	3	32	86 (65–97)	97 (84–100)
Akamatsu et al 2010 [12]	CT	6	3	0	13	100 (54–100)	81 (54–96)
Endo et al 2007 [20]	CT	4	4	0	7	100 (40–100)	64 (31–89)
Feydy et al 1999 [21]	CT	1	3	3	4	25 (0–81)	57 (18–90)
Watadani et al 2008 [19]	CT	0	0	0	13		
Summary estimates	CT					84 (63–94)	93 (69–99)
Looser et al 1992 [7]	Ultrasound	0	0	1	21	0	100 (84–100)
Neumaier et al 1995 [10]	Ultrasound	3	0	4	16	43 (10–82)	100 (79–100)
Nodal status							
Engels et al 1989 [25]	CT	17	1	0	4	80 (28–100)	80 (28–100)
Lee et al 2006 [22]	CT	8	2	7	38	53 (27–79)	95 (83–99)
Unno et al 2007 [23]	CT	5	0	9	10	36 (13–65)	100 (69–100)
Akamatsu et al 2010 [12]	CT	3	4	3	12	50 (12–88)	75 (48–93)
Watadani et al 2008 [19]	CT	2	2	4	5	33 (4–78)	71 (29–96)
Summary estimates	CT					61 (28–86)	88 (74–95)
Li et al 2008 [11]	PET/CT	5	1	7	4	42 (15–72)	80 (28–100)
Metastatasis							
Akamatsu et al 2010 [12]	CT	3	4	3	12	50 (12–88)	75 (48–93)
Li et al 2008 [11]	PET/CT	5	4	1	7	56 (21–86)	88 (47–100)

CI, confidence interval; PET, positron emission tomography.

Lymph node metastasis

5 studies including 136 patients reported on sensitivity and specificity of CT for lymph node status evaluated by CT (Table 7). Summary estimates were 61% for sensitivity (95% CI, 28–86%) and 88% for specificity (95% CI, 74–95%).

One study, including only 17 patients, reported on the performance of PET/CT in evaluating nodal status [11]. The authors found a sensitivity and specificity for nodal status of 42% and 80%, respectively.

No studies were identified that evaluated the use of MRI or ultrasound for assessment of lymph node metastasis.

Distant metastasis

Only one study reported on the sensitivity and specificity of CT on metastasis, which were 67% (4/6) and 94% (15/16), respectively [12]. Yet these results must be interpreted with caution, since this study was published in 1989 and it has several serious methodological limitations, as demonstrated by the methodological quality assessment (Table 3).

One study reported on the performance of PET/CT in identifying distant metastasis [11], and found a sensitivity and specificity of 56% and 88%, respectively.

Head-to-head comparison

Two studies compared MRI with CT head to head [5, 6]. In one study, accuracies for ductal extent of the tumour for CT and MRI were 64% (9/14) and 71% (10/14), respectively [5]. In the second study, accuracies for ductal extent of the tumour were 80% (16/20) and 75% (15/20) for CT and MRI, respectively [6]. No statistical tests were executed owing to the very small patient numbers. Both studies only reported the accuracy for ductal extent of the tumour, and unfortunately did not report on the other staging parameters (vascular involvement or lymph node metastases).

Discussion

In this review of all imaging studies concerning the staging of hilar cholangiocarcinoma, we found only 16 studies (primarily concerning CT only), including in total 448 patients. Although our aim was to compare the four individual investigations, this was not feasible because of the insufficient availability of adequate studies on MRI, ultrasound and PET/CT. Only data on CT were sufficient for pooling the findings. Pooled accuracy of CT for assessment of ductal extent of the tumour was 86%. Pooled sensitivity and specificity of CT were 89% and 92% for assessment of portal vein involvement, 84% and 93% for assessment of hepatic artery involvement, and 61% and 88% for assessment of nodal status, respectively.

The results of this systematic review should be interpreted with caution because of several limitations. Firstly, the included studies have limited methodological quality, as was detected by using the QUADAS tool (Table 3). Moreover, probably most importantly, the time between the index test and reference standard was mentioned in only 9 studies, and was less than 31 days in 7 studies. Therefore, misclassification due to progression of disease may have occurred [3]. Results of studies reporting on diagnostic performance are hard to interpret without details on methodology, and consequently many

QUADAS items were scored as unclear. New diagnostic studies should be reported, adhering to the Standards for Reporting of Diagnostic Accuracy (STARD) [13]. The STARD initiative provides a checklist of items that should be included in the report of a study on diagnostic accuracy.

Secondly, a meta-analysis was not feasible for ultrasound, MRI and PET/CT owing to the low number of data sets and low number of patients in the data sets. Consequently, no comparisons could be made. Thirdly, heterogeneity of the studies was substantial, owing to differences in imaging technique (*e.g.* for CT varying between 4- and 64-slice CT); study quality; small sample sizes in the studies (and consequent variability as a result of chance); and patient population (*e.g.* differences in disease stage). To compensate for this problem, we performed random models to adjust for the heterogeneity. Fourthly, we found only two studies, including only 34 patients in total, that compared two imaging investigations in the same population [5, 6]. Ideally the diagnostic accuracy of competing imaging investigations is assessed in the same patient population. This enables a more accurate assessment of differences in investigations, and also should identify pros and cons more easily. Fifthly, another limitation could be publication bias. Because of the small number of studies, as well as small number of patients within the studies, we believe a funnel plot to investigate publication bias was not meaningful.

20 articles were excluded because no 2×2 contingency table could be extracted. An additional 22 of the 70 articles selected for full manuscript review were excluded because no specific hilar cholangiocarcinoma data could be retrieved. Often data regarding hilar cholangiocarcinoma patients are reported in combination with the findings in other bile duct cancers. Presumably, this is a result of the rareness of the disease, and consequent small patient numbers. From a molecular and cell biological point of view, the discussion is still ongoing as to whether bile duct cancers can be seen as one entity or should be separated according to their location (intrahepatic, hilar and distal) [14, 15]. Yet, from a radiological point of view, interpretation of vascular involvement is clearly different when assessing an intrahepatic tumour from when assessing a hilar tumour that grows directly adjacent to the vessels. Moreover, gallbladder cancer, which is known for its high likelihood of metastatic disease, is also frequently included in studies on bile duct cancer, which can significantly impede correct interpretation of results. Therefore, we believe that future diagnostic studies on bile duct cancer should separately report the specific hilar cholangiocarcinoma data in the publication.

In addition to the imaging investigations evaluated in this study, also more invasive techniques—such as endoscopic retrograde cholangiopancreatography, percutaneous transhepatic cholangiography, staging laparoscopy, endoscopic ultrasound, intraductal ultrasound and choledochoscopy—are used in some centres to improve staging of hilar cholangiocarcinoma. Yet, since the staging process starts with non-invasive imaging, we intentionally have not evaluated these techniques in this systematic review, and have focused on non-invasive imaging. Nonetheless, the staging accuracy for hilar cholangiocarcinoma patients could be improved, and exploratory laparotomies for unresectable patients could potentially be decreased by using these techniques.

Finally, ultrasound, CT, MRI and PET/CT are used nowadays alone or in various combinations with each other for staging of hilar cholangiocarcinoma, and for the choice for surgical resection with curative intent or a palliative treatment. As Table 4 clearly shows, evidence on the staging accuracy of these investigations is limited, especially for MRI, ultrasound and PET/CT. Moreover, no adequate head-to-head comparative studies exist. As a result, an accurate comparison of the investigations and an evidence-based guideline cannot be made yet. Thus, it is of vital importance that future studies will provide this evidence. These studies ideally should have a prospective design, although because of the time required for prospective studies on this rare disease, well-designed retrospective studies could also be of importance. Furthermore, most evidence is available about the use of CT for staging of hilar cholangiocarcinoma, yet these studies have serious methodological flaws and small patient numbers. However— probably even more importantly—no criteria for involvement of adjacent structures (*e.g.* of the portal vein) exist. For example, for pancreatic cancer several criteria for portal vein involvement, including the presence of tumour involvement exceeding 180° of the circumference of the portal vein, have been established [16, 17]. Hence, future CT studies on hilar cholangiocarcinoma should focus on radiological criteria that can accurately predict vessel involvement using contemporary scanners.

In conclusion, diagnostic accuracy studies of CT, MRI, ultrasound or PET/CT for staging of hilar cholangiocarcinoma patients are sparse, often with a low number of patients giving moderate methodological quality. Therefore, there is a need for new methodologically solid studies. Owing to the lack of evidence, an adequate comparison of the various investigations was not feasible. Most evidence is available regarding staging with CT, which seems to have acceptable accuracy for assessment of ductal extent, portal vein and hepatic artery involvement (sensitivity and specificity ranging between 84% and 90%). The sensitivity of CT for nodal status, however, seems to be low (61%).

References

1. Kloek JJ, van Delden OM, Erdogan D, ten Kate FJ, Rauws EA, Busch OR, et al. Differentiation of malignant and benign proximal bile duct strictures: the diagnostic dilemma. World J Gastroenterol 2008;14:5032–8.
2. Ito F, Cho CS, Rikkers LF, Weber SM. Hilar cholangiocarcinoma: current management. Ann Surg 2009;250:210–18.
3. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003;3:25.
4. Bismuth H, Nakache R, Diamond T. Management strategies in resection for hilar cholangiocarcinoma. Ann Surg 1992; 215:31–8.
5. Cho ES, Park MS, Yu JS, Kim MJ, Kim KW. Biliary ductal involvement of hilar cholangiocarcinoma: multidetector computed tomography versus magnetic resonance cholangiography. J Comput Assist Tomogr 2007;31:72–8.
6. Kim HM, Park JY, Kim KS, Park MS, Kim MJ, Park YN, et al. Intraductal ultrasonography combined with percutaneous transhepatic cholangioscopy for the preoperative evaluation of longitudinal tumor extent in hilar cholangiocarcinoma. J Gastroenterol Hepatol 2010;25:286–92.
7. Looser C, Stain SC, Baer HU, Triller J, Blumgart LH. Staging of hilar cholangiocarcinoma by ultrasound and duplex sonography: a comparison with angiography and operative findings. Br J Radiol 1992;65:871–7.
8. Hann LE, Greatrex KV, Bach AM, Fong Y, Blumgart LH. Cholangiocarcinoma at the hepatic hilus: sonographic findings. AJR Am J Roentgenol 1997;168:985–9.
9. Masselli G, Manfredi R, Vecchioli A, Gualdi G. MR imaging and MR cholangiopancreatography in the preoperative evaluation of hilar cholangiocarcinoma: correlation with surgical and pathologic findings. Eur Radiol 2008;18:2213–21.
10. Neumaier CE, Bertolotto M, Perrone R, Martinoli C, Loria F, Silvestri E. Staging of hilar cholangiocarcinoma with ultrasound. J Clin Ultrasound 1995;23:173–8.
11. Li J, Kuehl H, Grabellus F, Muller SP, Radunz S, Antoch G, et al. Preoperative assessment of hilar cholangiocarcinoma by dual-modality PET/CT. J Surg Oncol 2008;98:438–43.
12. Akamatsu N, Sugawara Y, Osada H, Okada T, Itoyama S, Komagome M, et al. Diagnostic accuracy of multidetector-row computed tomography for hilar cholangiocarcinoma. J Gastroenterol Hepatol 2010;25:731–7.
13. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. Radiology 2003;226:24–8.
14. Hezel AF, Deshpande V, Zhu AX. Genetics of biliary tract cancers and emerging targeted therapies. J Clin Oncol 2010;28:3531–40.
15. Jarnagin WR, Klimstra DS, Hezel M, Gonen M, Fong Y, Roggin K, et al. Differential cell cycle-regulatory protein expression in biliary tract adenocarcinoma: correlation with anatomic site, pathologic variables, and clinical outcome. J Clin Oncol 2006;24:1152–60.
16. Lu DS, Reber HA, Krasny RM, Kadell BM, Sayre J. Local staging of pancreatic cancer: criteria for unresectability of major vessels as revealed by pancreatic-phase, thin-section helical CT. AJR Am J Roentgenol 1997;168:1439–43.
17. Phoa SS, Tilleman EH, van Delden OM, Bossuyt PM, Gouma DJ, Lameris JS. Value of CT criteria in predicting survival in patients with potentially resectable pancreatic head carcinoma. J Surg Oncol 2005;91:33–40.
18. Chen HW, Lai EC, Pan AZ, Chen T, Liao S, Lau WY. Preoperative assessment and staging of hilar cholangiocarcinoma with 16-multidetector computed tomography cholangiography and angiography. Hepatogastroenterology 2009;56:578–83.
19. Watadani T, Akahane M, Yoshikawa T, Ohtomo K. Preoperative assessment of hilar cholangiocarcinoma using multidetector-row CT: correlation with histopathological findings. Radiat Med 2008;26:402–7.
20. Endo I, Shimada H, Sugita M, Fujii Y, Morioka D, Takeda K, et al. Role of three-dimensional imaging in operative planning for hilar cholangiocarcinoma. Surgery 2007;142:666–75.
21. Feydy A, Vilgrain V, Denys A, Sibert A, Belghiti J, Vullierme MP, et al. Helical CT assessment in hilar cholangiocarcinoma: correlation with surgical and pathologic findings. AJR Am J Roentgenol 1999;172:73–7.
22. Lee HY, Kim SH, Lee JM, Kim SW, Jang JY, Han JK, et al. Preoperative assessment of resectability of hepatic hilar cholangiocarcinoma: combined CT and cholangiography with revised criteria. Radiology 2006;239:113–21.
23. Unno M, Okumoto T, Katayose Y, Rikiyama T, Sato A, Motoi F, et al. Preoperative assessment of hilar cholangiocarcinoma by multidetector row computed tomography. J Hepatobiliary Pancreat Surg 2007;14:434–40.
24. Sugiura T, Nishio H, Nagino M, Senda Y, Ebata T, Yokoyama Y, et al. Value of multidetector-row computed tomography in diagnosis of portal vein invasion by perihilar cholangiocarcinoma. World J Surg 2008;32:1478–84.
25. Engels JT, Balfe DM, Lee JK. Biliary carcinoma: CT evaluation of extrahepatic spread. Radiology 1989;172:35–40.