



# Diagnostic accuracy of fetal MRI to detect cleft palate: a meta-analysis

Hanneke E. M. van der Hoek-Snieders<sup>1</sup> · Antonius J. M. L. van den Heuvel<sup>1</sup> · Harmieke van Os-Medendorp<sup>1,2</sup> · Digna M. A. Kamalski<sup>3</sup>

Received: 8 June 2018 / Revised: 17 July 2019 / Accepted: 9 October 2019 / Published online: 3 December 2019  
© The Author(s) 2019

## Abstract

This systematic review aims to determine the diagnostic accuracy of fetal MRI for detecting cleft palate in fetuses at risk for orofacial clefts. Pubmed, Embase, and CINAHL were searched systematically. A diagnostic study was included if it performed MRI (index test) and postnatal examination (reference test) in fetuses at risk for orofacial clefts. Methodological quality was assessed using the QUADAS-2. A meta-analysis was performed with a random-effects model, calculating the pooled sensitivity, specificity, and area under the curve. The search resulted in eight studies (334 fetuses) to be included: four prospective and four retrospective studies. The applicability concern was low. There was, however, a risk of selection and information bias. All studies showed that MRI well predicted the chance of cleft palate. The sensitivity results were homogeneous, but heterogeneity was assumed regarding the specificity estimate (Cochrane's  $Q$  test:  $p = 0.00$ ). The pooled sensitivity was 0.97 (95% CI 0.93–0.99); the pooled specificity was 0.94 (0.89–0.97). The area under the curve was 0.98 (95% CI 0.98–0.99).

**Conclusion:** This meta-analysis shows that MRI has an excellent sensitivity and good to excellent specificity for diagnosing cleft palate in fetuses at risk for orofacial clefts. Future research should assess applicability for clinical care.

## What is Known:

- Using ultrasound for prenatal detection of cleft palate leads to misdiagnosis frequently.
- MRI could potentially improve the prenatal detection rate of cleft palate.

## What is New:

- Eight studies describe the diagnostic accuracy of MRI for detecting cleft palate.
- Combined results show excellent sensitivity and good to excellent specificity.

**Keywords** MRI · Diagnostic accuracy · Cleft palate · Prenatal

## Abbreviations

AUC	Area under the curve	CL	Cleft lip
CI	Confidence interval	CLP	Cleft lip and palate
		CP	Cleft palate

Communicated by Piet Leroy

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s00431-019-03500-x>) contains supplementary material, which is available to authorized users.

✉ Digna M. A. Kamalski  
ent-research@umcutrecht.nl; D.M.A.Kamalski@umcutrecht.nl

Hanneke E. M. van der Hoek-Snieders  
hannekesnieders@gmail.com

Antonius J. M. L. van den Heuvel  
ajmlheuvel@gmail.com

Harmieke van Os-Medendorp  
h.vanos@saxion.nl

<sup>1</sup> Clinical Health Sciences, Faculty of Medicine, Utrecht University, Utrecht, The Netherlands

<sup>2</sup> Department of Dermatology/Allergology, University Medical Centre Utrecht, Utrecht, The Netherlands

<sup>3</sup> Department of Otorhinolaryngology Head and Neck Surgery, University Medical Centre Utrecht, Heidelberglaan 100, G05.129, 3584, CX Utrecht, The Netherlands

FP	False positive
FN	False negative
MRI	Magnetic resonance imaging
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies-2
sROC	Summary receiver operating characteristics curve
TP	True positive
TN	True negative
US	Ultrasound

## Introduction

Orofacial clefts are one of the most common facial malformations [14]. The prevalence of orofacial clefts differs from 10.2 per 10,000 in the United States and Western Europe to 20.0 per 10,000 in Japan [44]. This condition can be either a cleft lip (CL), a cleft palate (CP), or both CL and CP (CLP) [26]. Orofacial clefts often result in feeding problems and difficulties in speech [18]. Also, many children with CLP or CP have chromosomal abnormalities, with the highest prevalence in children with isolated CP [39]. Approximately 50% of the isolated CP's are associated with other malformations, such as congenital heart defects or urinary tract defects [26].

Orofacial clefts are diagnosed by physical examination and visual inspection of the infant directly after birth [17]. To evaluate the palate, the mouth is inspected with a flashlight, while a spatula is used to depress the tongue.

Orofacial clefts are increasingly diagnosed prenatally during the ultrasound (US) fetal anomaly scan, often performed around the 20th week of the pregnancy [14]. After detecting abnormalities during this US screening, it is advised to refer the pregnant women with the fetus to a tertiary center, where the diagnosis is further explored [26]. Fetuses with a family risk factor for CL, CP, or CLP can be referred directly to these centers, since orofacial clefts are strongly genetically related [6].

Prenatal diagnosis provides opportunities for prenatal counselling and planning postnatal management [12, 34]. Prenatal counselling aims to help parents to cope with having a baby with a cleft and allows parents to make extra preparatory arrangements for the baby. Prenatal counselling also includes giving parents expert advice and support in the choice of the feeding method (breast milk or formula) allowing parents to make the feeding decision prior to the birth [13]. A study including 29 couples reported that 96% of the parents considered prenatal counselling as beneficial [38]. Early detection of orofacial clefts can also enable organization of the delivery in a tertiary center [15] and lead to fewer medical complications [20].

Unfortunately, many clefts are not detected during the US screening, since the detection rate differs from 9 to 50% and is further limited by factors such as maternal obesity [1, 25]. Specifically the palate can often not be adequately visualized, due to acoustic shadowing, which results in misdiagnoses of the palate [1, 7]. A systematic review investigated the diagnostic accuracy of second-trimester US in low- and high-risk populations [25]. A total of 27 articles were included, among which six studies with fetuses at high risk for orofacial clefts. Most studies reported detection rates between 9 and 50%, but results varied widely, depending on the population (high/low risk), type of cleft, and gestational age at the time of the scan. Detection rates were lower in the high-risk population: 0 to 44% of the orofacial clefts were detected prenatally using US in fetuses at risk. For isolated CP, the detection rate was 0% in this population. Assessing the palate during prenatal screening is essential in fetuses at risk for orofacial clefts, since CP has the highest association with feeding difficulties and other abnormalities [13]. The presence of a cleft palate can sometimes be used to rule out other diagnoses, such as a tumor, an encephalocele, or a vascular of lymphatic malformation [34]. Also, a recent systematic review investigated the prevalence of associated anomalies related to cleft category and concluded that prenatal counselling should be tailored to cleft category (CL, CP, or CLP) [26]. Without accurate information about the condition of the palate, this recommendation cannot be fulfilled.

Studies show magnetic resonance imaging (MRI) has potential to improve the prenatal detection rate of CP and should be considered as a complement to the current diagnostic procedure in fetuses at risk for a CP [31, 35]. Fetuses are at risk if the US screening shows abnormalities or if a family risk factor is present [2]. To evaluate the diagnostic value of MRI, information about the sensitivity and specificity of performing prenatal MRI for detecting CP is needed. A high sensitivity of the test is important to provide the prenatal counselling, including information about specific risks of CP. On the other hand, a high specificity prevents needless diagnoses, subsequent stress, and further medical investigations that could be a risk for the fetus. Namely, invasive genetic tests are indicated for CLP and CP with or without associated anomalies, but not for CL [26]. Therefore, the authors consider sensitivity and specificity of equal importance in the prenatal diagnosis of CP.

To decide if the prenatal diagnostic procedure for fetuses at risk for orofacial clefts should be extended by performing an MRI, information about its diagnostic value is needed. A few studies with relative small sample sizes were conducted to estimate the sensitivity and specificity of MRI for detecting CP [13, 27, 28]. Currently, no systematic review or meta-analysis has been conducted. Therefore, the diagnostic accuracy of MRI in detecting CP remains unknown. The aim of this meta-analysis is to establish the sensitivity and specificity of MRI in the prenatal diagnosis of cleft palate in fetuses at risk for orofacial clefts.

## Methods

### Design

To establish the diagnostic accuracy of MRI for the prenatal detection of CP, a meta-analysis was conducted according to the method of the Cochrane Handbook for Systematic Reviews and following the steps of the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement [16, 19].

### Information sources

The databases PubMed, Embase, and CINAHL were searched, because the subject of this meta-analysis is biomedical and related to the nursing field. Moreover, these databases all contain diagnostic studies, suitable for the research question. In case of good but insufficient data presented in the studies, authors were mailed to gain more specific results that could be used in the meta-analysis.

### Search method

Two reviewers (HH and AH) independently searched in the electronic databases on the 15th of June 2019. The search query combined the keywords fetuses, magnetic resonance imaging, cleft palate, and derivatives of these terms. The search strategy is provided in detail [Appendix](#). The keywords sensitivity and specificity were not included in the search query, because consistent use of those terms in electronic databases is doubted and could consequently lead to missing relevant articles [23]. The references of all included studies were examined.

### Inclusion criteria

The inclusion criteria included specifications of the patient-group and results of prior testing, since test accuracy can vary due to these factors [23]. A paper was included if (1) it was a full-text article, (2) published in English or Dutch, (3) included fetal MRI as index test and postnatal diagnosis as reference standard performed by physical examination or autopsy, (4) described the outcome as type of orofacial cleft with at least a distinction in CL and CP, (5) was conducted in a population of fetuses at risk for orofacial clefts (due to a positive or unclear US screening or orofacial cleft in a first degree relative), and (6) sensitivity or specificity were mentioned or calculable from study results.

Exclusion criteria were formulated concerning the research design, number of participants, and purpose of the study. Studies with < 10 participants were excluded, because small sample sizes cause concern regarding the power and precision of studies [5]. Case reports, editorials, opinions, dissertations,

intervention studies, and reviews were excluded as well as studies using fetal MRI for another purpose than diagnosing orofacial clefts. There was no restriction on the date of publication.

All included articles were also included in the meta-analysis, since the inclusion criteria were narrow regarding the population, index and reference test and calculability of sensitivity and specificity. A subgroup of the study sample was included if part of the sample did not meet the inclusion criteria.

### Study selection

After removing duplicates, the two reviewers (HH and AH) independently evaluated the eligibility of the selected studies from the literature in two phases. The first phase consisted of title and abstract screening and the second phase consisted of screening full-text articles. Disagreements were resolved in a consensus meeting.

### Data extraction

Data were extracted regarding study methodology, nation, patient population, inclusion and exclusion criteria, details about the performance of the index test, and details about the performance of the reference test. Outcome data on CP detected by MRI and the postnatal diagnosis were extracted in  $2 \times 2$  contingency tables, describing true positive (TP), false positive (FP), false negative (FN), and true negative (TN) results. Data extraction was performed independently by the two reviewers (HH and AH). Discrepancies between the reviewers were resolved in a consensus meeting.

### Data synthesis and statistical analysis

The statistical analysis was performed using Review Manager 5.3 and Meta-Disc software version 1.4 [40, 46]. Sensitivity and specificity along with 95% confidence intervals (CIs) were calculated for each study and presented in forest plots.

To determine the appropriateness of statistical pooling of accuracy estimates, homogeneity needed to be assessed [46]. Homogeneity means that the variation in the estimates of the test accuracy from different studies can be explained by study sampling error alone [9]. For this meta-analysis, the homogeneity of the sensitivity and specificity of the included studies was assessed in three ways [46]. First, the accuracy estimates of the forest plots were inspected visually. Estimates that lie along a line corresponding to the pooled accuracy estimate indicate homogeneity. Large deviations from this line indicate possible heterogeneity. Second, the Cochrane's  $Q$  test for heterogeneity was performed to evaluate if the differences across the studies are greater than expected by chance alone, using a chi-square ( $\chi^2$ ) distribution and  $k-1$  degrees of freedom. A  $p$  value < 0.05 suggests the presence of heterogeneity. Third, the inconsistency index I-squared was calculated to quantify the

amount of heterogeneity. An I-square > 50% indicates heterogeneity. Heterogeneity can arise from differences in cut-off point for defining a positive or negative result: a threshold effect [32]. Assessing a threshold effect is not applicable for this meta-analysis, because the thresholds used in interpreting MRI are unquantifiable since they depend on interpretation. Furthermore, a summary receiver operating characteristics curve (sROC) was presented to visualize the test sensitivity against the test specificity. Also, 95% CI was given. The area under the curve (AUC) was calculated and interpreted as follows: 0.5–0.6 = fail, 0.6–0.7 = poor, 0.7–0.8 = fair, 0.8–0.9 = good, 0.9–1 = excellent [29].

Finally, a weighted, pooled estimate of sensitivity and specificity was calculated using the random-effects method [19]. The weighting of the estimate depends on the number of participants included in this meta-analysis and is adjusted for the extent of heterogeneity.

### Methodological quality

The methodological quality of the included studies was assessed with the Quality Assessment of Diagnostic Accuracy Studies-2 (QUADAS-2), as the subject of this meta-analysis is determining diagnostic accuracy [45]. This instrument distinguishes between bias and applicability and consists of four domains: patient selection, index test, reference standard, and flow and timing. For risk of bias questions, the answer yes indicates high risk of bias, no indicates low risk of bias and the answer is “unclear” if insufficient data is reported to permit a judgment. Concerns regarding the applicability of results were raised if the patient selection, use of the index test, or the use of the reference test did not fully agree with the inclusion criteria of this meta-analysis.

Overall quality scores were not indicated, because different shortcomings may lead to different magnitudes of bias and it is difficult to fairly weigh each quality item [23]. Quality assessment was performed independently by the two reviewers (HH and AH) and all disagreements were resolved by consensus through discussion.

## Results

### Study selection

The search strategy yielded a total of 246 unique records. Furthermore, two additional results were found with reference checking. After screening on title and abstract, 19 articles appeared to meet the inclusion criteria. After reading the full-text articles, 11 studies were eventually excluded, resulting in a total of eight definite inclusions with a combined total of 334 fetuses [3, 11, 13, 21, 27, 28, 43, 47]. The flow diagram displaying exact details is presented in Fig. 1.

### Study characteristics

Study characteristics are shown in Table 1. The studies are all conducted in different countries: five countries in Europe [3, 11, 13, 27], one in North America [21], and two in Asia [43, 47]. All studies were published in 2010 or later. Four studies had a prospective design [11, 13, 28, 43] and four studies had a retrospective design [3, 21, 27, 47]. The size of the study populations varied from 12 to 94 fetuses. The mean gestation time was 27.59 weeks at the time of the MRI. All studies included participants based on a positive US screening. Three studies also included fetuses based on a family risk factor: first degree family member having CL, CP, or CLP [11, 13, 43]. Two studies excluded fetuses with associated abnormalities, such as agenesis of corpus callosum (a brain abnormality) [11, 28]. Seven studies used a 1.5-T MRI system as the index test of different brands. In one study, the type of MRI scanner was not specified [21]. Also, all studies used the postnatal diagnosis as the reference test performed by physical examination or autopsy. Autopsy was used in case of fetal death or termination of pregnancy.

### Methodological quality

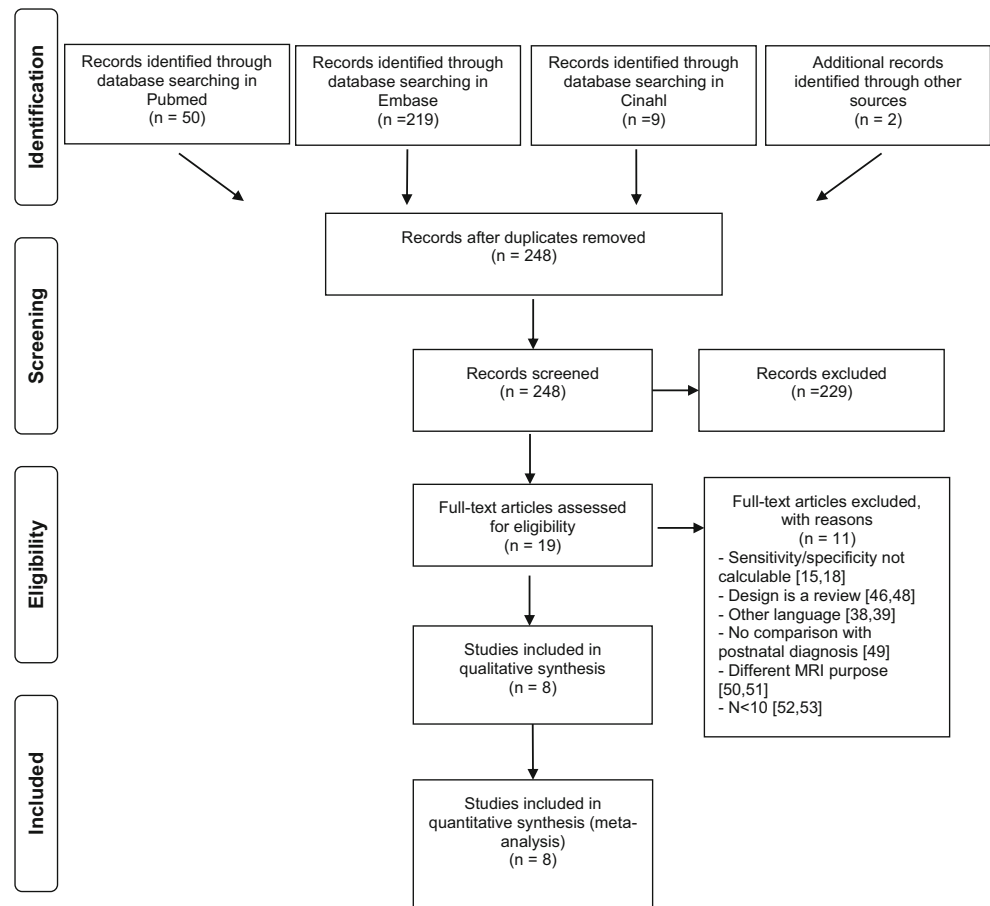
The risk of bias and applicability concerns are summarized in Table 2. The patient selection of four studies was judged as high risk of bias, since a case control design was not avoided and it was unclear whether a consecutive sample of patients was enrolled [3, 21, 27, 47]. In three studies, it was unclear how the participants enrolled [13, 28, 43]. One study clearly stated that all patients with CL, CLP, or CP were approached and included in the study after referral to two centers for further diagnostics [11].

Two studies described a blinded procedure for interpretation of the index test [11, 13]. The other studies were unclear about this procedure [1, 3, 21, 27, 43, 47]. Although technical specifications of the MRI scanner were frequently provided, the radiologists criteria for diagnosing CP were unclear in all studies.

Regarding the applicability of the studies, all studies had the main aim of determining the diagnostic accuracy of the MRI in detecting CP or CLP [3, 11, 13, 21, 27, 28, 43, 47]. All studies used postnatal examination as the reference test, by physical examination or by autopsy. One study did not report about using a reference standard, but the affirmation with the postnatal diagnosis was confirmed after contact with the author [3].

### Results of individual studies

The diagnostic accuracy of MRI in detecting CP is evaluated in this study in terms of the sensitivity and specificity. The results are visualized in Fig. 2 and Table 3.

**Fig. 1** Flow scheme of included studies

A combined total of 6 false negative diagnoses were reported in three studies [13, 21, 47] resulting in a sensitivity of respectively 0.93, 0.97, and 0.92 [13, 21, 47]. The other five studies had a sensitivity of 1.00 [3, 11, 27, 28, 43]. All sensitivity confidence intervals were in the range from 0.66 to 1.00.

A total of 10 fetuses received a false positive diagnosis. These diagnoses were made in four different studies and resulted in a specificity of respectively 0.76, 0.50, 0.92, and 0.88 [11, 13, 43, 47]. The other studies reported no false positive results, corresponding to a specificity of 1.00 [3, 21, 27, 28]. The specificity confidence intervals ranged from 0.01 to 1.00. The lowest specificity was reported in a study that included two negative cases, from which one case was correctly predicted [43].

### Homogeneity

The tabular results of the homogeneity analysis are included in Fig. 2. The forest plot shows that the estimates of the sensitivity lie roughly along the line of the pooled sensitivity (0.966). The Cochrane's  $Q$  test statistic is 6.63,  $df = 7$ ,  $p = 0.47$ , and the I-square is 0.0%. All three results indicate homogeneity.

For specificity, the forest plot shows more variation. Not all estimates lie along the line of the pooled specificity (0.94).

The Cochrane's  $Q$  test statistic is 22.82,  $df = 7$ ,  $p = 0.00$ , and the I-square is 69.3%. All specificity results indicate heterogeneity.

### Meta-analysis

The weighted pooled sensitivity from all eight studies was 0.97 (95% CI 0.928–0.988) and is interpreted as excellent. The weighted pooled specificity was 0.94 (0.89–0.97) and is interpreted as good to excellent. The sROC is provided in Fig. 3. The AUC is 0.98 (95% CI 0.98–0.99). This is interpreted as excellent.

### Discussion

This meta-analysis brings together all available evidence about the diagnostic accuracy of MRI for the prenatal diagnosis of CP. Combining the results of eight studies enabled to more precisely estimate the diagnostic accuracy of MRI, revealing excellent sensitivity and good to excellent specificity.

The methodological quality of this study is strengthened by the fact that the data collection, extraction, and synthesis were performed separately by two researchers. Also, the broad

**Table 1** Characteristics of included studies

First author	Year of publication	Nation	Design	No. of fetuses	Risk factor determined by	Exclusion criteria	Gestational time in weeks	Index test	Reference test
Bekicinska-Figatowska	2014	Poland	Retrospective	62	Positive US diagnosis	Not mentioned	Not mentioned	1.5 Tesla MRI	Postnatal findings
Dabadie	2016	France	Prospective	22	Positive US diagnosis or family risk factor	Associated abnormalities	29.5 (27–34)	1.5 Tesla MRI	Postnatal findings
Descamps	2010	United Kingdom	Prospective	49	Positive US diagnosis or family risk factor	Not mentioned	34.4 (24–37)	1.5 Tesla MRI	Postnatal findings
Latifer-Narin	2019	United States	Retrospective	61*	Positive US diagnosis	No reference test available	26.4 (18–38)	Not mentioned	Postnatal- or autopsy findings
Mallath-Pokorny	2010	Austria	Retrospective	34	Positive US diagnosis	Not mentioned	26.0 (19–34)	1.5 Tesla MRI	Postnatal- or autopsy findings
Manganaro	2011	Italy	Prospective	27**	Positive US diagnosis	Associated anomalies	23.7 (19–33)	1.5 Tesla MRI	Postnatal- or autopsy findings
Wang	2011	Japan	Prospective	12	Positive US diagnosis or family risk factor	Not mentioned	28.0 (21–34)	1.5 Tesla MRI	Postnatal- or autopsy findings
Zheng	2019	China	Retrospective	94***	Positive US diagnosis	Not mentioned	26.1 (19–38)	1.5 Tesla MRI	Postnatal- or autopsy findings

No, number; NA, not applicable; US, ultrasound; MRI, magnetic resonance imaging

\*19/61 Fetuses were lost to follow up or underwent termination of pregnancy without autopsy and not included in the meta-analysis

\*\*2/27 Fetuses in twin pregnancy were not at risk for CP and not included in the meta-analysis

\*\*\*6/94 Fetuses were lost to follow-up and not included in the meta-analysis

**Table 2** Summary of quality assessment using QUADAS-2

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Bekiesinska-Figatowska 2014	-	?	?	?	+	+	+
Dabadie 2016	+	?	?	+	+	+	+
Descamps 2010	?	?	?	+	+	+	+
Laifer-Narin 2019	-	?	?	-	+	+	+
Mailath-Pokorny 2010	-	-	-	+	+	+	+
Manganaro 2011	?	?	?	+	+	+	+
Wang 2011	?	-	-	+	+	+	+
Zheng 2019	-	?	?	-	+	+	+

+, low risk; -, high risk; ?, unclear risk

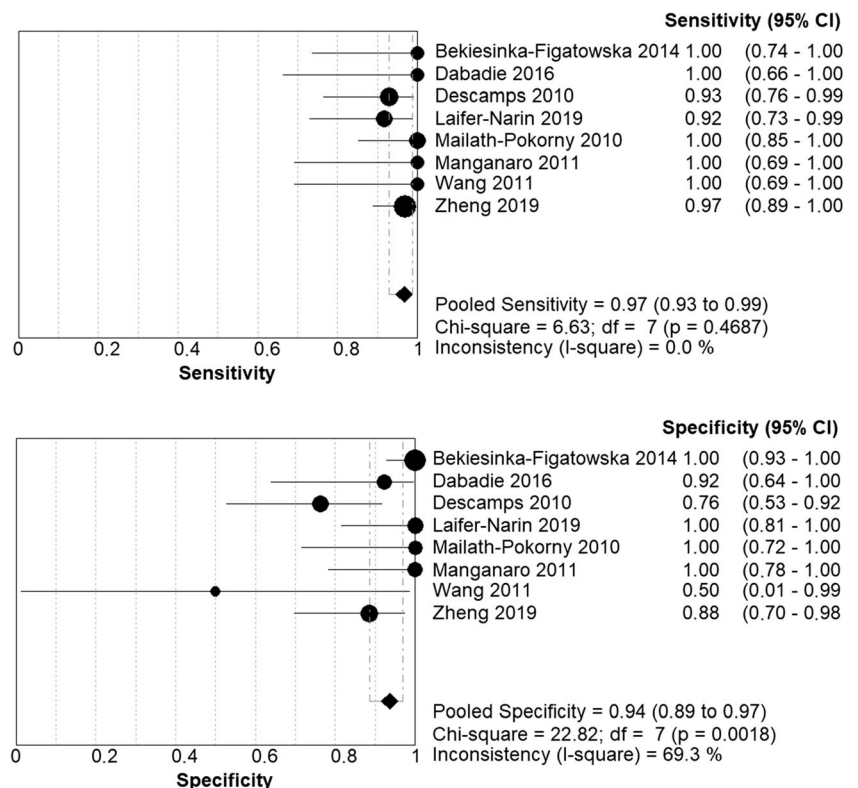
inclusion criteria used in this study, without additional search filters, contribute to the discovery of almost all available studies [23]. Furthermore, the power and precision of this study were increased, because a meta-analysis was performed [10]. However, some limitations are worth noting.

The reliability of this study was limited, because a small number of studies was included of which most included a relatively small sample size. Also, this meta-analysis does not provide information for subgroup analysis within the selected population. For instance, comparison could not be made between the diagnostic accuracy of isolated CP with

CLP, because isolated CP was described in only five fetuses [3, 11, 27]. One study did not provide information about the total number of isolated CP and CLP diagnosis [13]. Consequently, the false negative diagnosis and two false positive diagnosis that occurred in this study could not be further analyzed. The other two false positive diagnosis occurred two fetuses with CLP [11, 43]. The impact of gestational age could also not be explored, because it was not mentioned in one study [3].

Furthermore, selection bias might have been caused by the language filter that was used. The search strategy revealed two

**Fig. 2** Forest plot of all included studies. CI: confidence interval; TP: true positive; FP: false positive; FN: false negative; TN: true negative



**Table 3** Distribution of positive and negative diagnoses per study and whether these were predicted correctly with MRI

	True positive	False positive	False negative	True negative
Bekiesinska-Figatowska 2014	12	0	0	50
Dabadie 2016	9	1	0	12
Descamps 2010	26	5	2	16
Laifer-Narin 2019	22	0	2	18
Mailath-Pokorny 2010	23	0	0	11
Manganaro 2011	10	0	0	15
Wang 2011	10	1	0	1
Zheng 2019	60	3	2	23

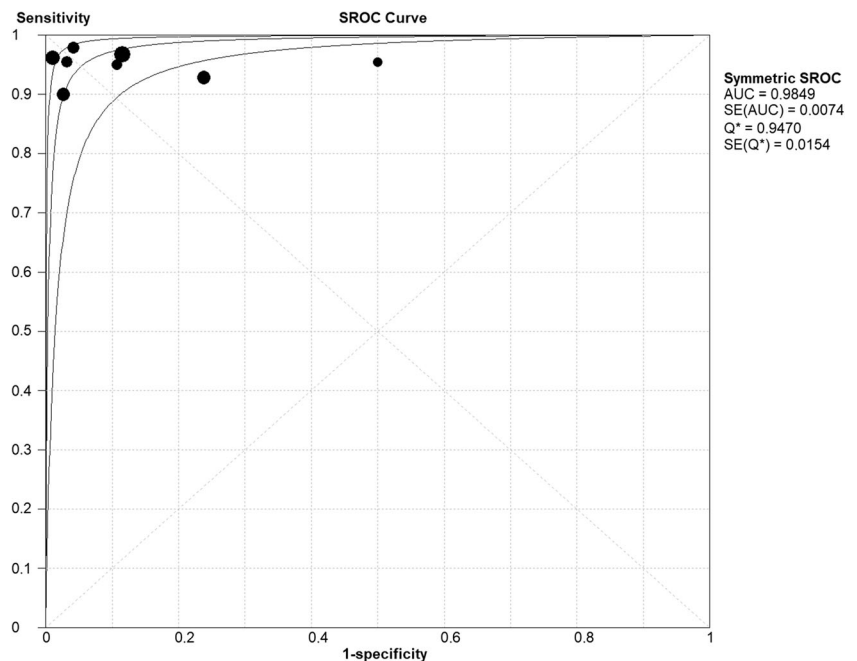
studies in other languages, one in Czech and one in Chinese [4, 42]. The language criterion might have caused even more selection bias if other studies did not publish an abstract in English. Additionally, information bias might be present, since the MRI specifications have not been included in the meta-analysis. Although all studies use a 1.5-Tesla MRI, the resolutions of these MRI's differ. The MRI resolution is likely to influence the diagnostic accuracy, but future research should confirm if a higher resolution contributes to a more accurate diagnosis.

Another limitation concerns the nature of the heterogeneity. Although the extent of heterogeneity was investigated thoroughly in this study, an assessment of the nature of the heterogeneity was not included in this meta-analysis for two reasons. First, a threshold effect, the primary cause for heterogeneity in diagnostic studies cannot be assessed, because the threshold in MRI assessment is implicit [23, 46]. It is therefore unclear to what extent study variation is caused by the application of different thresholds. Second, insufficient information is available about population variables and study characteristics to perform meta-

regression. For example, the time of the physical examination could have influenced the performance of the reference test and consequently could have caused variation in the estimated diagnostic accuracy [17]. Since the included studies did not explicitly state the time of the physical examination, this variable could not be included in a meta-regression.

Further, the reliability of the reference test can be discussed [11, 17, 22, 37]. For diagnostic studies, it is crucial to have a reference test that accurately determines the presence or absence of the target condition [37]. Postnatal diagnosis is considered the reference standard in diagnosing CP [11, 41]. This is performed by thorough physical examination or autopsy. However, some requirements are needed to rely on this diagnosis. The World Health Organization states the importance of visual inspection with a light and with a tongue depressor as part of this physical examination [17]. In daily practice, if this recommendation is not satisfied, delayed diagnosis of CP is not uncommon [41]. The studies included in this meta-analysis do not describe the way of performing the postnatal diagnosis and the value of the reference standard could be

**Fig. 3** Summary receiver operating characteristic plot of all included studies. AUC: area under the curve; SE: standard error; Q: *Q*-statistic





cause for discussion. Additionally, it is suggested that by gaining more experience with interpreting MRI, interpretation skills significantly improve [13]. It is unclear to what extent a difference in experience was present in the included studies, and if this difference has led to variation between the studies.

After determining the diagnostic accuracy of a test, the applicability needs to be investigated [23]. This includes assessment of the safety of the test, the availability of materials and expertise, the increased burden of care and cost-effectiveness. First, the increased use of MRI in prenatal diagnoses warrants questions about the safety of this diagnostic instrument, including the possibility of developmental defects due to the exposure to the magnetic radiation [22]. Two reviews show no harmful effects of MRI performed at 1.5 Tesla and lower [24, 35]. A 10-year follow-up study including was cited [33]. This study included 74 pregnant woman and did not reveal significant differences in birth weight between children that were exposed and not exposed to prenatal MRI. However, to our knowledge, the follow-up results have not been published yet. Visualization of fetal structures is more difficult and involves more risks in the first trimester of the pregnancy [8, 24]. It is advised to avoid performing MRI before 13 weeks gestational age [36]. The studies included in this meta-analysis performed MRI between 18 and 38 weeks of gestational age, and therefore meet this criterion. Second, performing prenatal MRI requires specific material and expertise, such as knowledge about MRI safety and how to optimize fetal images [24]. Prenatal MRI is advised therefore to perform in multidisciplinary centers with fetal MRI expertise [24]. Consequently, the mothers of fetuses at risk for CP will generally need to travel further, to visit a center where the MRI can be performed. Also, the length and costs of the hospital visit will be increased due to this additional diagnostic test. By assessing the applicability of MRI in prenatal diagnoses of CP, alternative options should also be considered.

Just as interest grows in diagnosing CP using prenatal MRI, improving prenatal diagnosis of CP through US techniques also receives attention [15, 30]. For example, using clinical markers derived from 2D- and 3D sonography is suggested to improve correct diagnosis of CP in the first trimester [30] and a scoring system has been developed to help sonographers to correctly identify CP [15]. To summarize all US innovations is beyond the scope of this study, but is recommended to be further explored in future research. Cost-effective studies including factors such as availability of equipment and skills are required to determine if MRI should be implemented in the diagnostic process of CP. It is recommended to include subgroups of fetuses in the decision-making process of diagnostic instruments for standard care, such as fetuses with a family risk factor of CP, fetuses with positive US diagnosis, fetuses with or without cleft lip, and fetuses with or without other anomalies. It is therefore not yet recommended to include MRI in standard prenatal care, but MRI could be used in specific cases relying on correct diagnosis.

In conclusion, this meta-analysis shows that MRI is an instrument with excellent sensitivity and good to excellent specificity for diagnosing CP in fetuses at risk for orofacial clefts. Findings indicate that the beneficial effect of performing MRI in this population should not be underestimated, since it provides opportunities to optimize prenatal management and counselling and planning of delivery and postnatal management. Additional research of good methodological quality should assess applicability for clinical care.

**Authors' contributions** Hanneke EM van der Hoek-Snieders Msc: design of the work, data collection, data analysis and interpretation, drafting the article, final approval of the version to be published

Antonius JML van den Heuvel Msc: data collection, critical revision of the article, final approval of the version to be published

Harmieke van Os-Medendorp PhD: critical revision of the article, final approval of the version to be published

Digna MA Kamalski MD, PhD: critical revision of the article, final approval of the version to be published

**Funding information** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest

**Informed consent** Informed consent was not applicable for this study, since it is a review.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

1. Arangio P, Manganaro L, Pacifici A, Basile E, Cascone P (2013) Importance of fetal MRI in evaluation of craniofacial deformities. *J Craniofac Surg* 24:773–776
2. Beaty TH, Marazita ML, Leslie EJ (2016) Genetic factors influencing risk to orofacial clefts: today's challenges and tomorrow's opportunities. *F1000Research* 5:2800
3. Bekiesińska-Figatowska M, Brągoszewska H, Romaniuk-Doroszevska A, Duczkowska A, Jaczyńska R, Maciejewski TM (2014) The role of magnetic resonance imaging in the prenatal diagnosis of cleft lip and palate. *Dev Period Med* 18:27–32
4. Bělobrádek Z, Žižka J, Klzo L et al (2013) Utilizing magnetic resonance imaging in prenatal detection of cleft lip and palate clefts. *Ces Radiol* 67:209–213
5. Biau DJ, Kernéis S, Porcher R (2008) Statistics in brief: the importance of sample size in the planning and interpretation of medical research. *Clin Orthop Relat Res* 466:2282–2288
6. Burg ML, Chai Y, Yao CA et al (2016) Epidemiology, etiology, and treatment of isolated cleft palate. *Front Physiol* 67:1–16

7. Burnell L, Verchere C, Pugash D et al (2014) Additional post-natal diagnoses following antenatal diagnosis of isolated cleft lip +/- palate. *Arch Dis Child Fetal Neonatal Ed* 99:286–290
8. Chen MM, Coakley FV, Kaimal A, Laros RKJ (2008) Guidelines for computed tomography and magnetic resonance imaging use during pregnancy and lactation. *Obstet Gynecol* 112:333–340
9. Chess LE, Gagnier JJ (2016) Applicable or non-applicable: investigations of clinical heterogeneity in systematic reviews. *BMC Med Res Methodol* 16:121–128
10. Collins R, Gray R, Godwin J, Peto R (1987) Avoidance of large biases and large random errors in the assessment of moderate treatment effects: the need for systematic overviews. *Stat Med* 6:245–250
11. Dabadie A, Quarello E, Degardin N, Desbriere R, Heckenroth H, Sigaudy S, Gorincour G (2016) Added value of MRI for the prenatal diagnosis of isolated orofacial clefts and comparison with ultrasound. *Diagn Interv Imaging* 97:915–921
12. Davalbhakta A, Hall PN (2000) The impact of antenatal diagnosis on the effectiveness and timing of counselling for cleft lip and palate. *Br J Plast Surg* 53:298–301
13. Descamps MJL, Golding SJ, Sibley J et al (2010) MRI for definitive in utero diagnosis of cleft palate: a useful adjunct to antenatal care? *Cleft Palate Craniofac J* 47:578–585
14. Ensing S, Kleinrouweler CE, Maas SM et al (2014) Influence of the 20-week anomaly scan on prenatal diagnosis and management of fetal facial clefts. *Ultrasound Obstet Gynecol* 44:154–159
15. Fuchs F, Burlat J, Grosjean F, Rayssiguier R, Captier G, Faure JM, Dumont C (2018) A score-based method for quality control of fetal hard palate assessment during routine second-trimester ultrasound examination. *Acta Obstet Gynecol Scand* 97:1300–1308
16. Guyatt GH, Oxman AD, Santesso N et al (2009) The Cochrane collaboration. *Review Manager (RevMan)*. *Pain* 9:205–213
17. Habel A, Elhadi N, Sommerlad B, Powell J (2006) Delayed detection of cleft palate: an audit of newborn examination. *Arch Dis Child* 91:238–240
18. Hanny KH, de Vries IAC, Haverkamp SJ, Oomen KP, Penris WM, Eijkemans MJ, Kon M, Mink van der Molen A, Breugem CC (2016) Late detection of cleft palate. *Eur J Pediatr* 175:71–80
19. Higgins J, Green S (editors) (2011) *Cochrane handbook for systematic reviews of interventions version 5.1.0 [updated March 2011]*. Cochrane Collab
20. Katorza E, Achiron R (2012) Early pregnancy scanning for fetal anomalies: the new standard? *Clin Obstet Gynecol* 55:199–216
21. Laifer-narin S, Schlechtweg K, Lee J et al (2019) A comparison of early versus late prenatal magnetic resonance imaging in the diagnosis of cleft palate. *Ann Plast Surg* 82:242–246
22. Lazarus E, Debenedectis C, North D, Spencer PK, Mayo-Smith WW (2009) Utilization of imaging in pregnant patients: 10-year review of 5270 examinations in 3285 patients-1997-2006. *Radiology* 251:517–524
23. Leeflang MMG, Deeks JJ, Gatsonis C, Bossuyt PMM (2008) Systematic reviews of diagnostic test accuracy. *Ann Intern Med* 149:889–897
24. Levine D (2013) Timing of MRI in pregnancy, repeat exams, access, and physician qualifications. *Semin Perinatol* 37:340–344
25. Maarse W, Bergé SJ, Pistorius L, van Bameveld T, Kon M, Breugem C, Mink van der Molen A (2010) Diagnostic accuracy of transabdominal ultrasound in detecting prenatal cleft lip and palate: a systematic review. *Ultrasound Obstet Gynecol* 35:495–502
26. Maarse W, Rozendaal AM, Pajkrt E, Vermeij-Keers C, Mink van der Molen A, van den Boogaard M (2012) A systematic review of associated structural and chromosomal defects in oral clefts: when is prenatal genetic analysis indicated? *J Med Genet* 49:490–498
27. Mailáth-Pokorny M, Worda C, Krampl-Bettelheim E, Watzinger F, Brugger PC, Prayer D (2010) What does magnetic resonance imaging add to the prenatal ultrasound diagnosis of facial clefts? *Ultrasound Obstet Gynecol* 36:445–451
28. Manganaro L, Tomei A, Fierro F et al (2011) Fetal MRI as a complement to US in the evaluation of cleft lip and palate. *Radiol Med* 116:1134–1148
29. Metz CE (1978) Basic principles of ROC analysis. *Semin Nucl Med* 8:283–298
30. Mookan S, Agnees J (2017) First-trimester sonographic evaluation of palatine clefts: a novel diagnostic approach. *J Ultrasound Med* 36:1397–1414
31. Moreira NC, Ribeiro V, Teixeira J et al (2013) Visualization of the fetal lip and palate: is brain-targeted MRI reliable? *Cleft Palate Craniofac J* 50:513–519
32. Moses LE, Shapiro D, Littenberg B (1993) Combining independent studies of a diagnostic test into a summary roc curve: data??analytic approaches and some additional considerations. *Stat Med* 12:1293–1316
33. Myers C, Duncan KR, Gowland PA, Johnson IR, Baker PN (1998) Failure to detect intrauterine growth restriction following in utero exposure to MRI. *Br J Radiol* 71:549–551
34. Nagarajan M, Sharbidre KG, Bhabad SH, Byrd SE (2018) MR imaging of the fetal face: comprehensive review. *RadioGraphics* 38:962–980
35. Pistorius LR, Hellmann PM, Visser GHA, Malinger G, Prayer D (2008) Fetal neuroimaging: ultrasound, MRI, or both? *Obstet Gynecol Surv* 63:733–745
36. Reddy UM, Filly RA, Copel JA (2008) Prenatal imaging: ultrasonography and magnetic resonance imaging. *Obstet Gynecol* 112:145–157
37. Reitsma JB, Rutjes AWS, Khan KS, Coomarasamy A, Bossuyt PM (2009) A review of solutions for diagnostic accuracy studies with an imperfect or missing reference standard. *J Clin Epidemiol* 62:797–806
38. Rey-Bellet C, Hohlfeld J (2004) Prenatal diagnosis of facial clefts: evaluation of a specialised counselling. *Swiss Med Wkly* 134:640–644
39. Shkoukani MA, Lawrence LA, Liebertz DJ, Svider PF (2014) Cleft palate: a clinical review. *Birth Defects Res C Embryo Today Rev* 102:333–342
40. The Nordic Cochrane Centre (2014) *Review manager*. Cochrane Collab 1–43
41. Van Veen-van der Hoek M, Jedeloo S, Bredero-Boelhouwer H et al (2014) Effects of delayed diagnosis in cleft palate: a preliminary qualitative study on the experience of parents. *Verpleegkd Spec* 9:30–35
42. Wang G-B, Chen L-G, Zhu X-Y et al (2010) MRI findings of fetal cleft lip and palate. *Chinese J Radiol* 44:152–155
43. Wang G, Shan R, Zhao L, Zhu X, Zhang X (2011) Fetal cleft lip with and without cleft palate: comparison between MR imaging and US for prenatal diagnosis. *Eur J Radiol* 79:437–442
44. Watkins SE, Meyer RE, Strauss RP, Aylsworth AS (2014) Classification, epidemiology, and genetics of orofacial clefts. *Clin Plast Surg* 41:149–163
45. Whiting P, Rutjes AWS, Reitsma JB et al (2003) The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol* 25:1–13
46. Zamora J, Abraira V, Muriel A et al (2006) Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 6:31–43
47. Zheng W, Li B, Zou Y, Lou F (2019) The prenatal diagnosis and classification of cleft palate: the role and value of magnetic resonance imaging. *Eur Radiol* 29:5600–5606. <https://doi.org/10.1007/s00330-019-06089-9>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.