Middle Ear Cholesteatoma: Non–Echo-planar Diffusion-weighted MR Imaging versus Delayed Gadolinium-enhanced T1-weighted MR Imaging—Value in Detection

Bert De Foer, MD
Jean-Philippe Vercruysse, MD
Anja Bernaerts, MD
Joke Meerschraet, MD
Christoph Kenis, MD
Marc Pouillon, MD
Luc De Beuckeleer, MD
Johan Michiels, PhD
Kris Bogaerts, PhD
Filip Deckers, MD
Thomas Somers, MD, PhD
Robert Hermans, MD, PhD
Erwin Offeciers, MD, PhD
Jan W. Casselman, MD, PhD

Purpose:
To retrospectively compare non–echo-planar (non-EP) diffusion-weighted (DW) imaging, delayed gadolinium-enhanced T1-weighted magnetic resonance (MR) imaging, and the combination of both techniques in the evaluation of patients with cholesteatoma.

Materials and Methods:
This institutional review board–approved study, for which the need to obtain informed consent was waived, included 57 patients clinically suspected of having a middle ear cholesteatoma without a history of surgery and 63 patients imaged before “second-look” surgery. Four blinded radiologists evaluated three sets of MR images: a set of delayed gadolinium-enhanced T1-weighted images, a set of non-EP DW images, and a set of both kinds of images. Overall sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV), as well as intra- and inter-observer agreement, were assessed and compared among methods. To correct for the correlation between different readings, a generalized estimating equations logistic regression model was fitted. Results were compared with surgical results, which were regarded as the standard of reference.

Results:
Sensitivity, specificity, NPV, and PPV were significantly different between the three methods (P < .005). Sensitivity and specificity, respectively, were 56.7% and 67.6% with the delayed gadolinium-enhanced T1-weighted images and 82.6% and 87.2% with the non-EP DW images. Sensitivity for the combination of both kinds of images was 84.2%, while specificity was 88.2%. The overall PPV was 88.0% for delayed gadolinium-enhanced T1-weighted images, 96.0% for non-EP DW images, and 96.3% for the combination of both kinds of images. The overall NPV was 27.0% for delayed gadolinium-enhanced T1-weighted images, 56.5% for non-EP DW images, and 59.6% for the combination of both kinds of images.

Conclusion:
MR imaging for detection of middle ear cholesteatoma can be performed by using non-EP DW imaging sequences alone. Use of the non-EP DW imaging sequence combined with a delayed gadolinium-enhanced T1-weighted sequence yielded no significant increases in sensitivity, specificity, NPV, or PPV over the use of the non-EP DW imaging sequence alone.

1 From the Department of Radiology (B.D.F., A.B., C.K., M.P., L.D.B., F.D., J.W.C.) and University Department of ENT (J.P.V., T.S., E.O.), Sint-Augustinus Hospital, Oosterweldlaan 24, 2610 Wilrijk, Belgium; Department of Radiology, Sint-Jan Hospital Antwerp, Brussels, Belgium (J.W.C.); Department of Radiology, University Hospitals Leuven, Leuven, Belgium (J. Meerschraet, R.H.); Interuniversity Institute for Biostatistics and Statistical Bioinformatics, Katholieke Universiteit Leuven, Leuven, Belgium and Université Hasselt, Hasselt, Belgium (K.B.); and Healthcare Sector, Siemens NV, Brussels, Belgium (J. Michiels). Received June 26, 2009; revision requested August 20; revision received October 27; accepted November 5; final version accepted December 10. Address correspondence to B.D.F.

*RSNA, 2010

In the past, computed tomography (CT) was considered the imaging technique of choice for the evaluation of middle ear cholesteatoma. Magnetic resonance (MR) imaging has gained increasing importance in the evaluation of the complicated middle ear cholesteatoma (1), in the postoperative follow-up of patients who have undergone middle ear surgery for cholesteatoma to detect residual cholesteatoma before “second-look” surgery (2–10), and in the evaluation of recurrent cholesteatoma (3,9–11).

Various MR imaging protocols have been proposed that are mainly based on the use of delayed gadolinium-enhanced T1-weighted sequences (2,4), diffusion-weighted (DW) imaging sequences (3,8,9,10), or a combination of both techniques (1,5–7,11). As regards DW imaging, distinction should be made between echo-planar (EP) DW imaging sequences (3,5,7,9,10) and non-EP DW imaging sequences (1,6,8,11). Non-EP DW imaging sequences have a thinner section thickness and a higher imaging matrix and are less prone to susceptibility artifacts than EP DW imaging sequences (6,8,12). The purpose of this study was to retrospectively compare non-EP DW imaging, delayed gadolinium-enhanced T1-weighted MR imaging, and the combination of both techniques in the evaluation of cholesteatoma.

Materials and Methods

Patients
This retrospective study was approved by the institutional review board of Sint-Augustinus Hospital, with a waiver of informed consent. We evaluated 120 patients with cholesteatoma (44 female patients [mean age, 35.8 years; range, 10–75 years] and 76 male patients [mean age, 35.4 years; range, 4–75 years]). The overall mean age was 35.7 years, with a range of 4–75 years. Data in patients before first-stage surgery and before second-look surgery were collected between July 1, 2005, and July 1, 2008. No MR imaging examination was degraded by motion artifacts, so no examination was excluded from the study. No significant difference in age distribution between sexes was present (P = .955). Patients undergoing first-stage surgery were similar in demographic variables, signs, symptoms, and disease status to patients undergoing second-look surgery.

Fifty-seven patients who were clinically suspected of having a middle ear cholesteatoma were included. They underwent an MR imaging study before first-stage surgery, which was performed within 2 weeks after imaging. The decision to perform first-stage surgery was made by the surgeon on the basis of clinical, otoscopic, audiologic, and CT findings.

Sixty-three patients who had undergone previous surgery for cholesteatoma (canal wall up tympanoplasty) were also included. They underwent an MR imaging study before second-look surgery that was designed to help search for residual cholesteatoma (21 patients) or recurrent cholesteatoma (42 patients). The decision to perform second-look surgery was made by the surgeon on the basis of findings at clinical follow-up and the findings at first-stage surgery. Surgery was performed within 2 months after imaging.

Imaging Technique
MR imaging was performed by using a 1.5-T superconducting unit (Magnetom Avanto; Siemens, Erlangen, Germany) with the standard head matrix coil and two 7-cm surface ring coils. To increase the signal-to-noise ratio, the 7-cm surface ring coils, which were receive only, were used together with the head matrix coil for all sequences except the whole-brain T2-weighted turbo spin-echo (SE) sequence. Axial 2-mm-thick SE T1-weighted images were obtained with the following parameters: repetition time msec/echo time msec, 400/17; matrix, 192 × 256; field of view, 150 × 200 mm; 12 sections; two acquisitions; acquisition time, 3 minutes 50 seconds. Coronal 2-mm-thick SE T1-weighted images were acquired with the same parameters except the matrix, which was set at 144 × 256.

Coronal 2-mm-thick turbo SE T2-weighted images (3500/92; matrix, 256 × 256; field of view, 250 × 250 mm; 12 sections; two acquisitions; acquisition time, 3 minutes 50 seconds; flip angle, 90°; repetition time msec/echo time msec, 2000/90) were acquired with the following parameters: repetition time msec/echo time msec, 5000/100; matrix, 192 × 256; field of view, 150 × 200 mm; 12 sections; two acquisitions; acquisition time, 4 minutes 20 seconds; flip angle, 90°; repetition time msec/echo time msec, 2000/90; matrix, 192 × 256; field of view, 150 × 200 mm; 12 sections; two acquisitions; acquisition time, 4 minutes 20 seconds; flip angle, 90°.

Abbreviations:

CI = confidence interval
DW = diffusion weighted
EP = echo planar
MPV = negative predictive value
PPV = positive predictive value
SE = spin echo

Author contributions: Guarantor of integrity of entire study, B.D.F.; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; manuscript final version approval, all authors; literature research, B.D.F., J.P.V.; clinical studies, B.D.F., L.P.V., A.B., J. Meersschaut, M.P., F.D., R.H., E.O., J.W.C.; statistical analysis, B.D.F., K.B.; and manuscript editing, B.D.F., J.P.V., A.B., C.K., M.P., L.D.B., J. Michiels, K.B., F.D., R.H., E.O., J.W.C.

Authors stated no financial relationship to disclose.
192 x 256; field of view, 130 x 200; turbo factor, 13; 12 sections; two acquisitions; acquisition time, 2 minutes 41 seconds) and axial 0.4-mm-thick three-dimensional turbo SE T2-weighted images (1500/303; matrix, 228 x 448; field of view, 107 x 210 mm; turbo factor, 37; 48 sections; one acquisition; acquisition time, 6 minutes 19 seconds) were also obtained. In all patients, a 2-mm-thick single-shot turbo SE MR DW sequence was performed in the coronal plane (2000/115; matrix, 134 x 192; field of view, 220 x 220 mm; b factors, 0 and 1000 sec/mm²; 20 sections; six signals acquired; imaging time, 2 minutes 14 seconds).

All sequences were performed 45 minutes after intravenous injection of 0.1 mol gadoterate meglumine (Dotarem; Guerbet, Roissy, France) or gadopentetate dimeglumine (Magnevist; Schering, Berlin, Germany) per kilogram of body weight. MR imaging was performed by the same group of four technicians, each of whom had at least 8 years of experience in MR imaging.

### Image Evaluation

Three data sets were retrospectively evaluated. In the first data set, delayed gadolinium-enhanced T1-weighted standard MR images—that is, axial and coronal delayed gadolinium-enhanced SE T1-weighted images, coronal turbo SE T2-weighted images, and three-dimensional turbo SE T2-weighted images—were evaluated alone. In the second data set, non-EP DW imaging, the single-shot turbo SE DW images were evaluated alone. Apparent diffusion coefficient maps were not calculated. In a third data set, the combined data set, images obtained with all sequences were evaluated together. All images were made anonymous and put in a random order.

All data sets were analyzed twice by four radiologists, with an interval of at least 2 weeks between readings. Two radiologists had long-standing experience in head and neck radiology (observers 1 [A.B.] and 2 [R.H.], with 7 and 18 years of experience, respectively). The other two radiologists (observers 3 [J. Meersschaut, a resident in radiology] and 4 [M.P.]) had little or no experience in MR imaging of the head and neck or the middle ear. All radiologists were blinded to patient identity, clinical and surgical findings, and CT data.

Cholesteatoma was diagnosed if a marked hyperintensity in comparison with brain tissue was noted on DW images obtained with a b factor of 1000 sec/mm². Images obtained with standard MR imaging sequences were evaluated for a moderately hyperintense lesion on T2-weighted images and the characteristic peripheral enhancing cholesteatoma matrix and a central nonenhancing area on delayed gadolinium-enhanced T1-weighted images. All sets of images were classified as "definitely yes," "probably yes," "probably no," or "definitely no." Surgical findings were considered the standard of reference.

Cholesteatoma surgery was performed by one of two experienced surgeons (T.S. and E.O.). The surgeons had more than 20 (T.S.) and 30 (E.O.) years of experience in cholesteatoma surgery at a tertiary referral center. Surgical results were classified as indicating cholesteatoma, residual cholesteatoma, recurrent cholesteatoma, or no cholesteatoma. Histologic evaluation was not performed.

### Statistical Analysis

The results of the original classification into four classes were dichotomized into "no" and "yes" for the purpose of analysis by combining the "definite" and "probable" categories. For only the intraobserver agreement for the different methods, the four categories were also analyzed. Interobserver agreement for each method was estimated by using the κ coefficient of agreement. Intraobserver agreement for each combination of reader and method was calculated by means of (weighted) κ values. κ Values were interpreted as suggested by Landis and Koch (13).

The diagnostic accuracy of each of the three methods was described by sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV).

To correct for the correlation between the different readings, a generalized estimating equation logistic regression model that assumed an independent working correlation matrix was fitted. An interaction term between observer and method was tested. In the case of no interaction, the average diagnostic measure over observers for each method was calculated. A global test for differences between methods was performed, and post-hoc pairwise testing between the methods was corrected for multiple testing by using the Bonferroni-Holm method. A significance level of 0.05 was used, and a corresponding 95% confidence interval (CI) was always calculated. With the sample size of our study, reasonable 95% CI widths could be obtained.

For analysis of the interobserver agreement and diagnostic accuracy, the first reading of each observer was used. All analyses were performed with software (SAS, version 9.2; SAS Institute, Cary, NC).

### Results

At surgery, 95 (79.2%) cholesteatomas were found in a total of 120 patients. Fifty (88%) cholesteatomas were found in the 57 patients in the first-stage group; 15 (30%) were regarded as small and/or empty retraction pockets. Forty-five (71%) cholesteatomas were found in the 63 patients in the second-look group; 11 (24%) were considered residual cholesteatomas and 34 (76%) were considered recurrent cholesteatomas.

The overall κ coefficient of interobserver agreement showed substantial agreement for the non-EP DW images (κ = 0.788) and the combined images (κ = 0.781). The κ coefficient for the delayed gadolinium-enhanced T1-weighted images showed fair agreement (κ = 0.363) (Table 1).

Intraobserver reliability, measured with a weighted κ coefficient of agreement, showed almost perfect agreement for both the non-EP DW images and the combined images for the four observers (with κ values ranging from 0.814 to 0.982). For the delayed gadolinium-enhanced T1-weighted images, observers 1 and 3 showed almost perfect agreement (κ = 0.894 and 0.826, respectively).
and observers 2 and 4 showed moderate agreement (κ = 0.549 and 0.518, respectively) (Table 2).

A significant interaction between method and observer was found for sensitivity (P = .049) but not for specificity (P = .30). For ease of interpretation, the interaction for sensitivity was also dropped.

Overall sensitivity for the delayed gadolinium-enhanced T1-weighted images was 56.7% (95% CI: 49.2%, 63.8%), with a specificity of 67.6% (95% CI: 53.0%, 79.4%). Overall sensitivity for the non-EP DW images was 82.6% (95% CI: 74.8%, 88.3%), with a specificity of 87.2% (95% CI: 69.0%, 95.4%). Overall sensitivity for the combined data set was 84.2% (95% CI: 76.7%, 89.6%), with a specificity of 88.2% (95% CI: 70.7%, 95.8%) (Table 3). Detailed sensitivity and specificity values per method and observer can be found in Table E1 (online).

No statistically significant difference in sensitivity or specificity could be shown between the non-EP DW images and the combined images (P = .157 and P = .705, respectively), whereas there were statistically significant differences in sensitivity and specificity between the delayed gadolinium-enhanced T1-weighted images and the combined images (P < .001 and P = .004, respectively). There was also a significant difference in sensitivity and specificity between the delayed gadolinium-enhanced T1-weighted images and the non-EP DW images (P < .001 and P = .006, respectively).

A significant interaction between method and observer was found for NPV (P = .002) but not for PPV (P = .36). For ease of interpretation, overall NPVs were still calculated.

The overall PPV for the delayed gadolinium-enhanced T1-weighted images was 88.0%, that for the non-EP DW images was 96.0%, and that for the combined images was 96.3%. The NPV for the delayed gadolinium-enhanced T1-weighted images was 27.0%, that for the non-EP DW images was 56.5%, and that for the combined images was 59.6% (Table 4). Detailed PPV and NPV values according to method and observer can be found in Table E2 (online).

No statistically significant difference in PPV could be shown between the non-EP DW images and the combined images (P = .62), whereas there was a significant difference in PPV between the delayed gadolinium-enhanced T1-weighted images and the combined images (P = .003). There was also a significant difference in PPV between the delayed gadolinium-enhanced T1-weighted images and the non-EP DW images (P = .004). For each observer, a significant difference in PPV between methods was seen (P < .0026 for all observers). For all observers, no significant difference could be shown between the non-EP DW images and the combined images, but a significant difference was found between the delayed gadolinium-enhanced T1-weighted images and the combined images and...
This gave rise to various signal intensities, making it difficult to recognize the characteristic signal intensities of cholesteatomas. The strength of the non-EP DW imaging sequence is that only cholesteatoma shows high signal intensity on images obtained with a $b$ factor of 1000 sec/mm$^2$ (Fig 2).

Recent publications demonstrated that non-EP DW imaging sequences have higher sensitivity and specificity in the detection of residual middle ear cholesteatoma than do EP DW imaging sequences (8). This is explained by the fact that non-EP DW imaging sequences have thinner section thicknesses, higher spatial resolution, and a complete lack of susceptibility artifacts at the interface between the temporal lobe and temporal bone (12).

In an electronic letter, Williams (14) states that delayed gadolinium-enhanced T1-weighted imaging remains comparable results were obtained by the noneXperienced and experienced readers, especially for the non-EP DW images and the images obtained with the combined protocol. These results suggest that even noneXperienced readers are able to correctly diagnose the presence of a cholesteatoma by using the non-EP DW imaging sequence.

In the literature, the highest sensitivity, specificity, PPV, and NPV in the diagnosis of residual cholesteatoma have been reported for delayed gadolinium-enhanced T1-weighted sequences (2,4), as well as for the combination of delayed gadolinium-enhanced T1-weighted sequences and non-EP DW imaging sequences (6,8).

The lower diagnostic accuracy for delayed gadolinium-enhanced T1-weighted images can probably be explained by the fact that our patient population included patients being imaged before first-stage and second-look surgery, without selection on the basis of CT findings. This resulted in a mixture of completely aerated, partially aerated and/or opacified, and completely opacified middle ears, mastoids, mastoidectomy cavities, and resection cavities (Fig 1).
were too small to allow us to draw valid inferences. No formal power analysis was performed before the study was started. For the patient population at hand, however, the sample size is quite large compared with that in other studies (1–4,6–9,11).

A probable bias of our study might be the high number of cholesteatomas detected at surgery. In the first-stage surgery subgroup, this can be explained by the fact that patients were included on the basis of clinical suspicion. In the second-look surgery subgroup, this can be explained by the fact that in our institution, patients with negative MR imaging findings before second-look surgery do not undergo surgery. As we considered surgical results as the standard of reference, we included only patients undergoing more sensitive for small lesions than DW imaging and that both techniques are needed for the detection of cholesteatoma. The fact that in our study no significant difference in sensitivity, specificity, PPV, and NPV between non-EP DW imaging and the combined protocol could be found opens the possibility of performing MR imaging in patients with cholesteatoma by using the non-EPI sequence alone, avoiding the need for intravenous contrast agent administration.

The relatively low NPV in our study can be explained by the mixture of patients undergoing first-stage surgery with those undergoing second-look surgery and by the presence of a relatively high number of small and/or empty retraction pockets in the first-stage group. These findings are in line with those in the literature (1,5), in which small and/or empty retraction pockets are responsible for false-negative findings at diffusion-weighted imaging in patients before first-stage surgery.

On the basis of these results, our imaging protocol for cholesteatomas has been changed. Delayed gadolinium-enhanced T1-weighted sequences are no longer used. To better localize suspected lesions identified at non-EP DW imaging, an axial and coronal turbo SE T2-weighted sequence is added. This results in a substantial shortening of imaging time. This will also result in an important cost saving for the health care system and a higher patient throughput for MR imaging.

No subgroup analyses between the first-stage and second-look surgery groups were performed in our study because the respective sample sizes were too small to allow us to draw valid inferences. No formal power analysis was performed before the study was started. For the patient population at hand, however, the sample size is quite large compared with that in other studies (1–4,6–9,11).

A probable bias of our study might be the high number of cholesteatomas detected at surgery. In the first-stage surgery subgroup, this can be explained by the fact that patients were included on the basis of clinical suspicion. In the second-look surgery subgroup, this can be explained by the fact that in our institution, patients with negative MR imaging findings before second-look surgery do not undergo surgery. As we considered surgical results as the standard of reference, we included only patients undergoing
surgery, resulting in a relatively high number of cholesteatomas at second-look surgery. In this retrospective study, this high positive number is less relevant, as the goal of the study was not to evaluate the detection rate of cholesteatoma before second-look surgery but to compare different MR imaging techniques in patients with cholesteatoma.

Another cause of probable bias could be the delay between MR imaging and surgery, which was longer in the second-look surgery subgroup. This also seems less relevant, as cholesteatoma can be performed by patients suspected of having middle ear cholesteatoma with delayed contrast-enhanced MR imaging: initial findings. Eur Radiol 2003;13(1):169–174.


References
