

Endoscopy Outperforms CT and DWI-MRI in Detecting Recurrent or Residual Middle Ear Cholesteatoma: A 140-Patient Real-Life Study

ABSTRACT

Objective: To evaluate the "real-life" diagnostic accuracy of high definition otoendoscopy compared with computed tomography (CT) and magnetic resonance imaging (MRI) in detecting cholesteatoma recurrence, and to assess the impact of surgical experience and technique on diagnostic reliability.

Study Design: Retrospective clinical study.

Setting: Tertiary referral center.

Patients: 140 patients undergoing surgical revision for suspected recurrent cholesteatoma. All patients underwent preoperative otoendoscopy; a subgroup of patients also underwent CT (n=87), MRI (n=62), or both (n=35) based on clinical judgment.

Interventions: Preoperative otoendoscopic assessments (performed by a senior otologist and an ENT resident) and available imaging findings were compared with intraoperative results (Reference Standard).

Main Outcome Measures: Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and Area Under the Curve (AUC). Inter-observer agreement was assessed using Cohen's kappa (κ) and McNemar's test. Multivariable logistic regression was used to identify independent predictors of recurrence across different subgroups.

Results: Intraoperative recurrence was confirmed in 80% of cases. Senior otoendoscopy showed superior diagnostic performance (Sensitivity: 81.3%; PPV: 91.9%; AUC: 0.763) compared to CT (AUC: 0.633) and MRI (AUC: 0.570). Overall inter-observer agreement was substantial ($\kappa = 0.68$; $p < 0.001$). However, stratification by technique revealed that 89.5% of discrepancies occurred in Canal Wall Up (CWU) procedures, while agreement reached 93.3% in Canal Wall Down (CWD) cases. In multivariable models tailored to the subgroups with available imaging, otoendoscopy remained the only significant independent predictor of recurrence.

Conclusions: High definition otoendoscopy is a reliable, accessible, and highly accurate tool for postoperative cholesteatoma follow-up. While surgical experience is crucial in complex CWU cases, the technique's overall reproducibility is high. Given its superior performance in a real-life setting, systematic "blanket" MRI protocols for all postoperative patients appear unjustified. A clinically driven approach centered on otoendoscopy optimizes diagnostic accuracy and healthcare resource allocation.

INTRODUCTION

Cholesteatoma remains one of the most challenging chronic otologic conditions to diagnose and monitor after surgical treatment. Every surgical procedure carries a variable risk of recurrent or residual disease, and early identification of these conditions is essential to avoid complications and to reduce the need for more demolitive revision surgeries.

Until the early 2000s, second-look surgery represented the gold standard for detecting residual cholesteatoma in previously operated ears. The introduction of non-echo planar diffusion-weighted MRI (non-EPI DWI) transformed this paradigm, shifting second-look procedures from a mandatory routine to a targeted intervention. Numerous authors have proposed structured MRI follow-up protocols and highlighted the importance of repeated imaging to improve detection accuracy. Despite its excellent sensitivity and specificity for residual pearls, routine MRI surveillance for all postoperative cholesteatoma patients raises important concerns: MRI is expensive, time-consuming, and, in high-volume referral centers, not always feasible on a large scale.

In parallel, the introduction of otoendoscopy has revolutionized the otologic clinical exam. High-definition close-up visualization of the tympanic membrane, combined with the ability to record images and videos, has improved the clinician's ability to detect subtle anomalies and early keratin debris. In everyday practice, not every patient who has undergone cholesteatoma surgery requires MRI, and even when MRI is indicated, a rigid standardized protocol does not fit all cases. Multiple factors—including surgical technique, extension of the primary cholesteatoma, cavity anatomy, and individual healing patterns—directly influence the optimal follow-up strategy.

Despite this evolution, the available literature still focuses predominantly on radiologic methods—particularly MRI and CT—for diagnosing recurrent disease, whereas relatively few studies have directly compared the diagnostic performance of radiology with that of clinical

and otoendoscopic assessment, tools that are more readily available in routine outpatient practice. This gap is particularly relevant in “real-life” clinical settings, where resource allocation, operator experience, and patient-specific factors often shape decision-making.

Building on these considerations, the present study was designed to systematically compare the diagnostic accuracy of otoendoscopy, CT, and DWI-MRI—individually and in combination—in a real-life postoperative cohort undergoing evaluation before revision surgery. Additional aims were to assess whether examiner experience influences otoendoscopic diagnostic performance, to investigate the impact of otorrhea on endoscopic evaluation accuracy, and to determine whether combining multiple diagnostic tools enhances overall accuracy compared with the use of single modalities.

MATERIALS AND METHODS

A retrospective observational study was conducted at a tertiary referral center, examining postoperative patients evaluated during outpatient follow-up for suspected residual or recurrent cholesteatoma. All assessments were performed as part of routine clinical practice, reflecting real-life diagnostic conditions. The study was conducted in accordance with the Declaration of Helsinki. Given the retrospective nature of the study and the use of de-identified data collected during standard clinical care, formal approval from the local Ethics Committee was not required according to national regulations.

All patients who underwent cholesteatoma surgery between January 2018 and August 2025 in our department were identified.

Inclusion criteria were:

- prior surgical treatment (CWU, CWD, or endoscopic approach) in the same ear, performed either at our center or elsewhere;
- clinical and/or radiological suspicion of residual or recurrent cholesteatoma;
- availability of clinical reports (including symptoms, otorrhea specifically) and otoendoscopic videoclips;
- availability of imaging reports, if performed;
- availability of surgical records from the revision procedure;

Exclusion criteria were:

- first cholesteatoma surgery on a specific side;
- congenital cholesteatoma surgery.

For each patient, a comprehensive database was generated. Recorded variables included the side of the revision surgery and the type of the initial procedure (CWU or CWD). Otorrhea, defined as either observed during evaluation or reported by the patient, was treated as a binary variable (present vs. absent).

All patients underwent otoendoscopy using 3-mm 0° endoscopes paired with a Full-HD or 4K camera system. Each clip was independently reviewed by an ENT resident and a senior otologist who were blinded to the definitive surgical outcome. Reviewers were provided only with basic clinical information (side and type of previous surgery) to simulate real-world practice. Each reviewer indicated the presence or absence of residual cholesteatoma as a binary variable.

Imaging assessment was based on radiologists' reports. CT was considered positive if any of the following conditions were met: (1) the radiologist explicitly stated that the findings were compatible with cholesteatoma; (2) the presence of a well-defined, lobulated soft-tissue mass within the middle ear was reported; or (3) isodense material was described in association with clear erosion of middle-ear bony structures. MRI was considered positive only when the radiologist explicitly reported the presence of cholesteatoma within the ear. Radiologic evaluation was therefore treated as a dichotomous variable (positive vs negative). Consistent with established clinical protocols aimed at minimizing post-operative inflammatory artifacts and mitigating the risk of false negatives due to lesions below the MRI detection threshold, all MRI examinations in this cohort were performed at least 12 months after the primary surgical procedure¹.

For each patient who underwent revision surgery, the operative report was reviewed to determine whether residual cholesteatoma was present. This surgical finding was considered the reference standard (outcome measure) against which all clinical, otoendoscopic, and radiologic assessments were compared.

Statistical analyses were performed using Stata 18.0 (StataCorp LLC, College Station, TX, USA). Continuous variables were described as mean $\pm$ standard deviation (SD), while categorical variables were summarized as absolute frequencies and percentages. Diagnostic accuracy of

the different tests (otoendoscopy, computed tomography, magnetic resonance imaging) was assessed by calculating sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV), along with their corresponding 95% confidence intervals (95% CI).

Inter-observer agreement between the senior otologic surgeon and the ENT resident was evaluated using Cohen's κ coefficient and interpreted according to Landis and Koch's criteria: slight (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.00). Potential asymmetry in discordant cases was examined with McNemar's test. Associations between categorical variables were explored using Pearson's χ^2 test.

To identify independent predictors of recurrence, multivariable binary logistic regression models were constructed using a manual enter method. To account for missing data in imaging studies, comparative analyses were performed using sub-group specific models. Specifically, when comparing the diagnostic performance of different combinations (e.g., otoendoscopy + CT), only patients with available data for all variables in that specific model were included (e.g., $n=87$ for the CT sub-analysis), and the reference ROC curve was re-calculated on the same sub-sample to ensure internal validity.

Model discrimination was estimated using the Area Under the ROC Curve (AUC). All tests were two-tailed, and a p -value < 0.05 was considered statistically significant. Comparison between ROC curves derived from different multivariable models was performed using the DeLong test for paired data.

RESULTS

Initially, we identified a cohort of 460 patients for whom both complete preoperative clinical documentation and operative reports confirming a diagnosis of cholesteatoma were available. These cases, treated at our department between January 2018 and August 2025, served as the primary screening pool for this study. From this initial group, 230 patients were excluded as they had undergone primary surgery for acquired or congenital disease, leaving 230 cases of suspected residual or recurrent cholesteatoma for further selection.

Following an additional selection process, only patients with clinical suspicious for whom preoperative otoendoscopic video recordings and clinical documentation regarding symptoms—particularly the presence of otorrhea—were available were included in the study. A total of 140 patients met these criteria, while 90 were excluded due to incomplete data.

The characteristics of the study sample are summarized in Table 1.

Table 1 - Demographic, clinical, and diagnostic characteristics of the study population (N = 140). Abbreviations: SD, standard deviation; CWU, canal wall up; CWD, canal wall down; CT, computed tomography; MRI, magnetic resonance imaging.

Variable	Total Sample (n)	n (%) or Mean $\pm$ SD
Demographics		
Age (years)	140	35.87 $\pm$ 20.75
Sex (Male)	140	88 (62.9%)
Clinical Presentation		
Side (Right)	140	66 (47.1%)
Otorrhea (Present)	140	49 (35.0%)
Previous Surgery Type		
- CWU	140	110 (78.6%)
- CWD	140	30 (21.4%)
Diagnostic Test Findings		
Senior Otoendoscopy (Positive)	140	99 (70.7%)
ENT resident Otoendoscopy (Positive)	140	96 (68.6%)
Availability of Preoperative Imaging		
CT scan	140	87 (62.1%)
MRI	140	62 (44.3%)
CT scan + MRI	140	35 (25%)
Preoperative Imaging Findings		
CT scan (Positive)	87	49 (56.3%)
MRI (Positive)	62	41 (66.1%)
Intraoperative Finding (Reference Standard)		
Recurrence (Prevalence)	140	72 (51.4%)
Residual (Prevalence)	140	40 (28.6%)

Regarding MRI assessments, a significant heterogeneity in reporting was observed. In several instances, the specific sequences utilized were not clearly detailed, and Diffusion-Weighted Imaging (DWI) was not consistently mentioned in the radiologic reports. Furthermore, the majority of reports lacked a clear distinction between non-echo-planar (non-EPI) and echo-planar (EPI) sequences. It is noteworthy that these examinations were frequently performed and reported by general radiologists in various non-specialist centers, rather than dedicated neuroradiology units

Otoendoscopy – Senior Otologist assessment

The otoendoscopic assessment performed by the senior otologic surgeon, when compared with the intraoperative reference standard, demonstrated a sensitivity of 81.3% (95% CI: 72.8–88.0) and a specificity of 71.4% (95% CI: 51.3–86.8). The positive predictive value (PPV) was 91.9% (95% CI: 84.7–96.4), while the negative predictive value (NPV) was 48.8% (95% CI: 32.9–64.9). The corresponding contingency matrix was as follows: TP = 91, FP = 8, FN = 21, and TN = 20 (n = 140). The overall discriminative ability of otoendoscopic assessment, expressed as the area under the ROC curve (AUC), was 0.763 (95% CI: 0.671–0.856).

Otoendoscopy – ENT Resident assessment

The otoendoscopic assessment performed by the ENT resident otologic surgeon, when compared with the intraoperative reference standard, demonstrated a sensitivity of 77.7% (95% CI: 68.8–85.0) and a specificity of 67.9% (95% CI: 47.6–84.1). The positive predictive value (PPV) was 90.6% (95% CI: 82.9–95.6), while the negative predictive value (NPV) was 43.2% (95% CI: 28.3–59.0). The corresponding contingency matrix was as follows: TP = 87, FP = 9, FN = 25, and TN = 19 (n = 140). The overall discriminative ability of otoendoscopic assessment, expressed as the area under the ROC curve (AUC), was 0.728 (95% CI: 0.631–0.824).

Inter-observer Agreement: Senior Otologist vs. ENT Resident

Direct comparison between the otoendoscopic assessments performed by the senior otologist and the ENT resident showed an overall observed agreement of 86.4%. Cohen's κ was 0.68 (Standard Error: 0.084; $p < 0.001$), indicating substantial agreement according to Landis and Koch's criteria.

When analyzing positive and negative agreement separately, the positive agreement was 90.3%, while the negative agreement was 77.6%. Disagreement was observed in 19 cases (13.6%): 8 out of 41 cases (19.5%) classified as negative by the senior surgeon were judged

positive by the resident, while 11 out of 99 cases (11.1%) classified as positive by the senior surgeon were judged negative by the resident. McNemar's test on discordant pairs showed no significant asymmetry ($p \cong 0.6$), suggesting that there was no statistically significant systematic difference between the two reviewers in their overall tendency to classify cases as positive or negative.

To further investigate the impact of surgical anatomy on diagnostic reliability, the cohort was stratified by technique (Table 2). In the CWD subgroup ($n=30$), the observed agreement was high at 93.3% ($\kappa = 0.634$; $p < 0.001$), with only 2 discordant cases (6.7%) where the senior surgeon identified a recurrence not seen by the resident. McNemar's test showed no significant systematic asymmetry ($p = 0.157$).

In contrast, the CWU subgroup ($n=110$) showed a lower agreement of 84.6% ($\kappa = 0.664$; $p < 0.001$). Notably, 17 out of the 19 total discordant cases in the entire study (89.5%) occurred within this group. The disagreement was balanced, with 8 cases rated positive only by the resident and 9 cases only by the senior surgeon. McNemar's test confirmed the absence of systematic bias ($p = 0.808$), suggesting that the discrepancies in CWU cases reflect intrinsic diagnostic ambiguity rather than a consistent observer bias.

Table 2 - Inter-observer agreement (Senior vs Resident) stratified by surgical technique.

Technique	n	Agreement (%)	Cohen's κ (p)	Discordant Cases	McNemar (p)
Overall	140	86.4%	0.68 (<0.001)	19	0.600
CWU	110	84.6%	0.66 (<0.001)	17 (89.5%)	0.808
CWD	30	93.3%	0.63 (<0.001)	2 (10.5%)	0.157

Diagnostic Accuracy of CT

In the subgroup of patients with available CT imaging ($n = 87$), comparison with the intraoperative reference standard yielded a sensitivity of 60.0% (95% CI: 48.0–71.1) and a specificity of 66.7% (95% CI: 34.9–90.1). The positive predictive value (PPV) was 91.8% (95% CI: 80.4–97.7), while the negative predictive value (NPV) was 21.1% (95% CI: 9.55–37.3). The corresponding contingency matrix showed: True Positives (TP) = 45, False Positives (FP) = 4, False Negatives (FN) = 30, and True Negatives (TN) = 8 (cholesteatoma prevalence in this subsample: 86%).

The overall discriminative performance of CT, as measured by the area under the ROC curve (AUC), was 0.633 (95% CI: 0.483–0.783), indicating limited discrimination in distinguishing between patients with and without recurrence. In this clinical setting, characterized by a high prevalence of recurrence, the combination of moderate sensitivity and specificity resulted in a high PPV but a notably low NPV.

Diagnostic Accuracy of MRI

In the subgroup of patients with available MRI imaging (n = 62), comparison with the intraoperative reference standard demonstrated a sensitivity of 68.6% (95% CI: 54.1–80.9) and a specificity of 45.5% (95% CI: 16.7–76.6). The positive predictive value (PPV) was 85.4% (95% CI: 70.8–94.4), while the negative predictive value (NPV) was 23.8% (95% CI: 8.22–47.2). The corresponding contingency matrix showed: True Positives (TP) = 35, False Positives (FP) = 6, False Negatives (FN) = 16, and True Negatives (TN) = 5 (cholesteatoma prevalence in this subsample: 82%). Notably, all imaging studies were conducted with a minimum interval of 12 months from the previous surgery.

The overall discriminative performance of MRI, expressed by the area under the ROC curve (AUC), was 0.570 (95% CI: 0.403–0.738), indicating low discrimination in distinguishing between patients with and without recurrence. In this MRI subgroup, the combination of moderate sensitivity and low specificity resulted in a high PPV but a very low NPV, consistent with the high prevalence of recurrence within this sample. A comprehensive comparison of the diagnostic performance of all tested modalities is detailed in Table 3.

Table 3 - Diagnostic performance of Otoendoscopy (Senior and Resident), CT, and MRI.

Test	n	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	PPV	NPV	AUC
Otoendoscopy (S)	140	91	8	21	20	81.3% (72.8–88.0)	71.4% (51.3–86.8)	91.9%	48.8%	0.763
Otoendoscopy (R)	140	87	9	25	19	77.7% (68.8–85.0)	67.9% (47.6–84.1)	90.6%	43.2%	0.728
CT scan	87	45	4	30	8	60.0% (48.0–71.1)	66.7% (34.9–90.1)	91.8%	21.1%	0.633
MRI	62	35	6	16	5	68.6% (54.1–80.9)	45.5% (16.7–76.6)	85.4%	23.8%	0.570

Performance of Otoendoscopy in Relation to Otorrhea

Recurrence was more frequent in patients with otorrhea (87.8%; 43/49) compared to those without (75.8%; 69/91); however, the overall association between otorrhea and recurrence did not reach statistical significance ($\chi^2(1) = 2,83$; $p=0,092$).

When stratifying the senior surgeon's otoendoscopic performance by otorrhea status, the results were as follows:

- in patients without otorrhea ($n = 91$): TP = 51, FP = 4, FN = 18, TN = 18. This corresponds to a sensitivity of 73.9% (95% CI: 61.9–83.7), a specificity of 81.8% (59.7–94.8), a PPV of 92.7% (82.4–98.0), and an NPV of 50.0% (32.9–67.1);
- in patients with otorrhea ($n = 49$): TP = 40, FP = 4, FN = 3, TN = 2. This yielded a sensitivity of 93.0% (95% CI: 81.4–98.0), a specificity of 33.3% (9.7–70.0), a PPV of 90.9% (78.8–96.4), and an NPV of 40.0% (11.8–76.9).

The presence of otorrhea was associated with increased sensitivity of otoendoscopy at the expense of lower specificity. PPV remained high in both subgroups, while NPV remained modest.

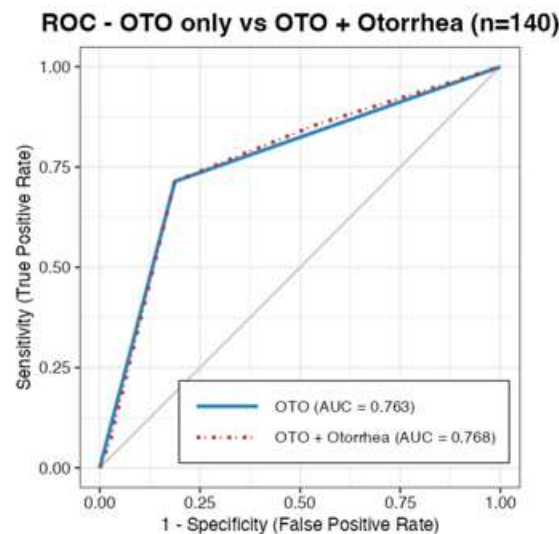
Multivariable Model: Otoendoscopy and Otorrhea

The results of the multivariable logistic regression models for predicting recurrence are summarized in Table 4. In the overall sample ($n = 140$), a multivariable logistic regression model including otoendoscopy (senior assessment) and otorrhea demonstrated good discriminative capacity (AUC = 0.768; Pseudo- $R^2 = 0.198$). Among the predictors, otoendoscopy was significantly associated with intraoperative recurrence ($\beta = 2.38$; SE = 0.52; $p < 0.001$; OR ≈ 10.8 ; 95% CI: 4.3–27.5), whereas the presence of otorrhea did not show a significant independent contribution ($\beta \approx 0$; $p \approx 0.99$). The inclusion of otorrhea in the model did not substantially improve overall discrimination compared to the model based on otoendoscopy alone (AUC = 0.763) (Figure 1). Although the inclusion of otorrhea in the model slightly increased the AUC from 0.763 to 0.768, this improvement was not statistically significant (DeLong test, $p = 0.857$), confirming that otorrhea does not add substantial discriminative power to the endoscopic assessment alone.

Table 4 - Multivariable logistic regression models for predicting intraoperative cholesteatoma recurrence. The table presents four distinct binomial logistic regression models. Model 1 includes the entire cohort (n=140), while Models 2, 3, and 4 refer to subgroups with available CT, MRI, or both, respectively. For all models, otoendoscopic assessment was based on the senior otologist's findings. Odds Ratios (OR) and their 95% Confidence Intervals (CI) are reported to represent the increased likelihood of recurrence associated with a positive diagnostic finding. Statistical significance was assessed using the Wald test for individual predictors, while model fit was estimated via McFadden's Pseudo-R² and Area Under the ROC Curve (AUC).

Model	Subgroup (n)	Predictor	Odds Ratio (95% CI)*	p-value	Pseudo-R²	AUC
1	Overall (140)	Otoendoscopy	10.8 (3.9–30.0)	< 0.001	0.198	0.768
		Otorrhea	1.00 (0.3–3.2)	0.997		
2	CT group (87)	Otoendoscopy	20.1 (3.5–114.3)	< 0.001	0.257	0.824
		Otorrhea	2.87 (0.5–16.6)	0.239		
		CT	2.66 (0.6–11.4)	0.182		
3	MRI group (62)	Otoendoscopy	7.9 (1.7–37.3)	0.009	0.176	0.787
		Otorrhea	0.67 (0.1–4.1)	0.667		
		MRI	2.15 (0.5–9.8)	0.321		
4	Combined (35)	Otoendoscopy + otorrhea + CT + MRI	<i>(Unstable Estimates)</i>	0.996	0.468	0.907

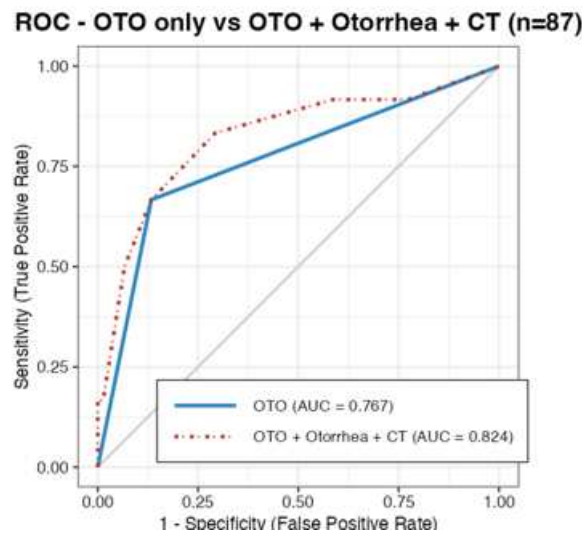
Figure 1. ROC curves illustrating the diagnostic performance of otoendoscopy alone (solid blue line) versus otoendoscopy combined with otorrhea (dotted red line) in the overall cohort (n=140). The addition of otorrhea did not significantly improve the AUC (p=0.857). (n=140). The addition of otorrhea did not significantly improve the AUC (p=0.857).



Multivariable Model: Otoendoscopy, Otorrhea, and CT

In the subgroup of patients with available CT data (n = 87), the logistic regression model including otoendoscopy (senior assessment), otorrhea, and CT findings demonstrated an AUC of 0.824 (Pseudo- $R^2 = 0.257$). Among the predictors, otoendoscopy remained the only statistically significant variable ($\beta = 2.999$; SE = 0.888; z = 3.38; p = 0.001; OR $\approx$ 20.1; 95% CI: 3.5–114.3). Neither otorrhea ($\beta = -1.055$; p = 0.239; OR $\approx$ 0.35) nor CT findings ($\beta = 0.980$; p = 0.182; OR $\approx$ 2.66) reached statistical significance as independent predictors. The overall discriminative capacity of this three-predictor model (AUC = 0.824) was moderately higher than that of the model based on otoendoscopy alone within the same subgroup (AUC = 0.767) (Figure 2). However, this difference did not reach statistical significance (DeLong test, p = 0.271), consistent with the logistic regression finding that CT was not an independent predictor of recurrence.

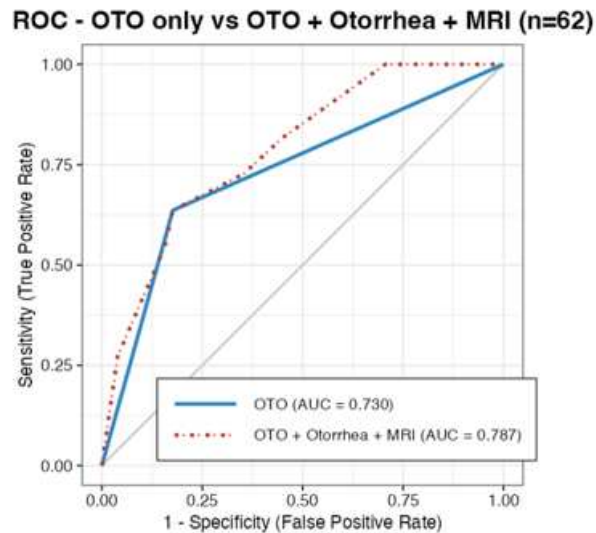
Figure 2. ROC curves comparing the diagnostic performance of otoendoscopy alone (solid blue line, recalculated for this subgroup) and otoendoscopy combined with otorrhea and CT findings (dotted red line) in the CT subgroup (n=87). The difference was not statistically significant (p=0.271).



Multivariable Model: Otoendoscopy, Otorrhea, and MRI

In the subgroup of patients with available MRI data (n = 62), the logistic regression model including otoendoscopy (senior assessment), otorrhea, and MRI findings demonstrated a fair overall discriminative capacity (AUC = 0.787; Pseudo-R² = 0.176). Among the predictors, only otoendoscopy was independently associated with recurrence ($\beta = 2.07$; SE = 0.79; z = 2.61; p = 0.009; OR $\approx$ 7.9; 95% CI: 1.7–37.3). Neither otorrhea ($\beta = 0.40$; p = 0.667; OR $\approx$ 1.5; 95% CI: 0.24–9.1) nor MRI findings ($\beta = 0.77$; p = 0.321; OR $\approx$ 2.1; 95% CI: 0.47–9.8) provided a significant independent contribution. The discriminative performance of this model (AUC = 0.787) was moderately higher than that of the model based on otoendoscopy alone (AUC = 0.730) (Figure 3). This improvement was not statistically significant (DeLong test, p = 0.134), suggesting that in this real-life cohort, MRI did not provide a statistically robust diagnostic gain over the endoscopic examination.

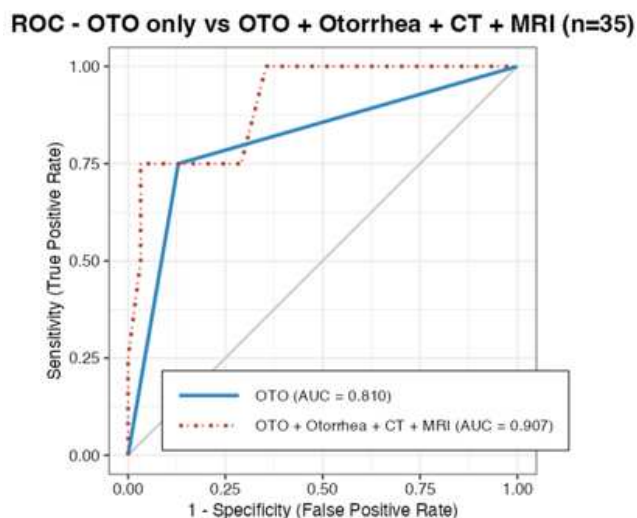
Figure 3. ROC curves comparing the diagnostic performance of otoendoscopy alone (solid blue line, recalculated for the MRI subgroup) and otoendoscopy combined with otorrhea and MRI findings (dotted red line) in patients with available MRI data (n=62). The difference was not statistically significant (p=0.134).



Multivariable Model: Otoendoscopy, Otorrhea, CT, and MRI

In the subgroup of patients with both imaging modalities available (n = 35), the logistic regression model including otoendoscopy (senior assessment), otorrhea, CT, and MRI findings demonstrated high discriminative capacity, with the AUC increasing from 0.810 (otoendoscopy alone) to 0.907 (combined model; Pseudo-R² = 0.468) (Figure 4). However, this difference did not reach statistical significance (DeLong test, p = 0.075), and none of the individual predictors reached significance in the regression model. Consequently, the high AUC value must be interpreted with caution and in a purely descriptive sense, as it may be influenced by the small sample size and potential overfitting.

Figure 4. ROC curves comparing the diagnostic performance of otoendoscopy alone (solid blue line, recalculated for the combined imaging subgroup) and otoendoscopy combined with otorrhea, CT, and MRI findings (dotted red line) in patients with both CT and MRI available (n=35). Although the combined model showed a higher AUC, the difference did not reach statistical significance (p=0.075), likely due to the small sample size.



DISCUSSION

The Persistent Challenge of Postoperative Cholesteatoma

Despite significant diagnostic and therapeutic advancements in recent decades, cholesteatoma remains a formidable challenge in otologic practice, particularly concerning postoperative follow-up. Notwithstanding continuous refinements in surgical techniques, recurrence rates remain substantial, ranging from 10% to 45% in adults and up to 40–50% in pediatric populations^{2,3}, possibly due to higher histological aggressiveness in children⁴. Incidence is notably higher after canal wall up procedures compared to canal wall down techniques⁵. Such figures underscore the vital necessity for accurate postoperative surveillance aimed at the early identification of residual or recurrent disease.

However, a universal consensus on the timing of follow-up and the criteria for surgical revision is still lacking, as these remain highly variable and heavily influenced by the experience of individual centers and surgeons. Especially after canal wall up procedures, diagnosing recurrence can be complex, leading many centers to adopt a mandatory second-look surgery at approximately 12 months⁶. While innovations like non-echo-planar diffusion-weighted MRI (non-EPI DWI) and image fusion techniques have improved detection capabilities, their application is not yet fully standardized. Our study was conceived to address this gap, evaluating the diagnostic accuracy of these tools within a real-life clinical setting.

The "High-Risk" Context of a Tertiary Referral Center

It is important to emphasize that our findings originate from a tertiary referral center specialized in complex otologic surgery. In such a setting, the inherent risk of residual or recurrent disease is naturally higher, as we frequently manage advanced cases or patients previously operated on in lower-level centers where surgical challenges might have been significant. This "high-risk" environment directly influences the diagnostic threshold and the clinical weight assigned to every suspicious finding.

Selection Bias and Prevalence

The high prevalence of recurrence observed in our study (80%) is a direct consequence of our stringent inclusion criteria. Unlike screening studies that evaluate all postoperative patients, our cohort was pre-selected based on a strong clinical or radiologic suspicion of residual or recurrent cholesteatoma. This targeted selection was designed to reflect the most difficult decision-making scenarios in clinical practice: those where a specialist must decide whether to return to the operating room. Therefore, the accuracy and predictive values reported here specifically describe the performance of diagnostic tools in a "high-suspicion" population rather than in a general follow-up setting.

The Economic Impact of Clinical Precision

In this context of high prevalence and clinical complexity, the "expert eye" during otoendoscopy provides a substantial economic advantage. Our data showed 16 cases where the radiologic report (MRI) was negative (False Negatives), yet the senior surgeon's suspicion led to a confirmed intraoperative finding of recurrence.

By proceeding directly to surgery based on otoendoscopic findings, we effectively bypassed the need for "watchful waiting" and the costs associated with serial imaging. Recent economic modeling suggests that while systematic MRI follow-up is generally more sustainable than mandatory second-look surgery, this advantage is strictly dependent on the recurrence rate. Specifically, when the prevalence of disease exceeds a certain threshold (calculated at 42%), a surgical approach becomes the superior strategy, offering better long-term quality-of-life outcomes by eliminating the risk of undetected disease⁷.

In our cohort, the 80% prevalence—resulting from a targeted clinical-radiologic selection—strongly supports an endoscopy-driven surgical intervention. Beyond immediate

expenditures, a reliance on imaging alone, even in specialized settings, carries a significant risk of missing residuals and recurrences due to the inherent sensitivity limits of MRI. Failing to identify these cases early—as seen in some of our patients with suspicious otoendoscopy but negative MRI reports—would inevitably lead to the management of more extensive disease. Such delays result in more demolitive procedures, higher complication rates, and increased long-term burdens on the healthcare system.

Furthermore, this clinical-driven approach avoids the cumulative costs of repeated MRI scans, which are often hampered by a 25% to 50% loss in patient follow-up over time, as reported in studies on serial imaging. While this logic is particularly evident in high-risk scenarios, it is even more pertinent for "low risk" cholesteatomas, such as those confined to the mesotympanum or epitympanum without medial ossicular chain involvement. In these cases, the diagnostic performance of high definition otoendoscopy could potentially replace routine MRI surveillance altogether. These considerations suggest that MRI should not be a mandatory tool for all postoperative patients but rather a targeted resource, prioritizing clinical and endoscopic assessment to optimize both patient outcomes and healthcare expenditure.

The Role of Otoendoscopy: Reliability across Experience Levels

A significant finding of this study is the diagnostic performance of otoendoscopy. The senior otologist achieved a sensitivity of 81.3% and a specificity of 71.4%, with a remarkably high PPV (91.9%). While the literature has predominantly focused on advanced radiologic methods (non-EPI DWI), few studies have quantitatively evaluated clinical outpatient tools. Our data suggest that otoendoscopy is a sensitive, immediate, and highly applicable instrument for follow-up.

Moreover, our analysis confirms the reproducibility of this method. Comparison with an ENT resident's assessments revealed an observed agreement of 86.4% and a Cohen's κ of 0.68 ($p < 0.001$), indicating substantial agreement. Accuracy values were comparable between the two observers: the senior surgeon showed a sensitivity of 81.3% and a specificity of 71.4%, while the resident obtained 77.7% and 67.9% respectively, with very similar PPV and NPV. The absence of significant systematic differences in the McNemar test further reinforces the role of otoendoscopy as a reliable tool even in less experienced hands, provided high-definition visualization is available.

Surgical Technique as a Diagnostic Modifier

However, our stratification analysis reveals that surgical experience becomes a decisive factor depending on the type of previous surgery. In CWD cavities, the agreement was nearly universal (93.3%). The "exteriorized" nature of these cavities provides a straightforward, panoramic view where recurrence is often obvious, effectively "leveling" the diagnostic performance between senior and junior evaluators.

Conversely, the CWU subgroup accounted for nearly 90% of all inter-observer discrepancies. This disparity is clinically significant: in CWU, the preservation of the posterior canal wall and the presence of a reconstructed tympanic membrane create anatomical recesses where residual disease can be easily overlooked or confused with post-operative scarring. The balanced nature of the disagreements in this group (McNemar $p = 0.808$) underscores that CWU procedures present an inherent diagnostic challenge. In these cases, the expert eye is essential not just to identify the lesion, but to correctly interpret subtle morphological changes that a junior otologist might either miss or over-interpret.

The Reality of Imaging: CT and MRI in Daily Practice

In our subgroup of 87 patients with available CT scans, diagnostic accuracy was moderate (S 60.0%, Sp 66.7%), with a very low NPV (21.1%). These results confirm the known limitations of CT in discriminating recurrence from inflammatory or post-surgical changes, where HRCT often shows sensitivities around 43% and specificities of 48% in the literature^{1,8-10}. Advanced techniques like Temporal Subtraction CT (TSCT) can increase sensitivity (up to 77% for experts), but they require standardized protocols and dedicated software not widely available⁸.

Similarly, for MRI (n=62), our results (S 68.6%, Sp 45.5%) were lower than the benchmarks reported in experimental studies, where non-EPI sequences achieve sensitivities between 91% and 100%¹¹⁻¹⁵. This discrepancy highlights the real-life nature of our study: our analysis was based on reports from various non-specialist centers where the specific DWI sequence (EPI vs. non-EPI) was often unspecified or suboptimal. While non-EPI DWI is technically superior, in a non-specialist setting, the radiologist's specific expertise remains crucial to distinguish cholesteatoma from inflammatory material¹⁶.

The Impact of Otorrhea as a Clinical Trigger

Otorrhea was analyzed as a primary clinical parameter because it is easily objectified, unlike hearing loss, which in post-surgical ears is often related to scarring or ossicular status rather

than recurrence. We observed that otorrhea acts as a diagnostic amplifier. In ears with otorrhea, otoendoscopic sensitivity increased to 93.0%, but specificity dropped to 33.3%, as secretions and granulation tissue visually amplify pathological signs, leading to overinterpretation.

Conversely, in dry ears, otoendoscopy proved more selective and reliable (Sp 81.8%), with a positive finding highly likely to represent a true recurrence. Literature confirms that while infection or secretion predisposes to recurrence¹⁷, its predictive value is limited¹⁸. Our findings suggest that while otorrhea makes the disease more visible, it reduces the discriminative capacity of the exam, necessitating integration with other modalities in wet ears.

Multivariable Integration and Discriminative Power

Multivariable analysis confirmed that otoendoscopy remains the most robust independent predictor of recurrence ($p < 0.01$; AUC ≈ 0.76). Integrating imaging modalities moderately improved the discriminative capacity: in the subgroup with CT, the AUC grew to 0.824, while in the MRI group, it reached 0.787. While imaging findings contributed positively, they did not reach independent statistical significance in the multivariable models, likely due to the superior predictive strength of the otoendoscopic exam and the aforementioned heterogeneity of radiologic reports.

In the small subgroup with both imaging modalities ($n = 35$), the combined model reached an excellent AUC of 0.907. However, this value must be interpreted with caution as a purely descriptive indicator. The limited sample size led to unstable coefficient estimates and potential overfitting, highlighting the need for larger, prospective cohorts to confirm the benefit of such an integrated approach.

Strengths and Limitations

The study's retrospective nature and the high prevalence of recurrence within the cohort (resulting in a limited number of True Negatives) may have influenced the stability of specificity and NPV estimates. Furthermore, the reliance on standard radiologic reports rather than a centralized review by dedicated neuroradiologists introduced unavoidable variability. However, these limitations paradoxically represent the study's primary strength: they faithfully reflect the complexity and variability of real-world clinical practice, providing a concrete perspective on the actual performance of these diagnostic tools in daily outpatient care.

CONCLUSIONS

The present study, conducted in a real-life clinical setting, evaluates the diagnostic accuracy of otoendoscopy, computed tomography, magnetic resonance imaging, and clinical assessment (otorrhea) in the diagnosis of cholesteatoma recurrence. Our findings highlight the central role of high definition otoendoscopy, which emerged as the most consistent and effective tool for discriminating recurrence. Notably, otoendoscopy demonstrated sensitivity and specificity levels superior to those of radiological methods in a real-world scenario, maintaining high inter-observer agreement regardless of the clinician's level of experience. This suggests that otoendoscopy is not only accurate but also highly reproducible and accessible, serving as a reliable reference method for routine outpatient surveillance.

While CT and MRI showed lower diagnostic performance in our cohort compared to the benchmarks reported in optimized experimental literature—likely due to the heterogeneity of reports and the lack of dedicated non-EPI DWI protocols in non-specialist centers—both modalities maintain an essential complementary role. Their value lies primarily in anatomical definition and preoperative surgical planning rather than as primary screening tools. Furthermore, while otorrhea was not found to be an independent predictor of recurrence, it acts as a clinical modifier that increases endoscopic sensitivity while potentially reducing specificity.

Importantly, our results suggest that the implementation of systematic, "blanket" MRI protocols for all postoperative patients is not justified, neither in terms of diagnostic efficacy nor healthcare resource allocation. Given the high positive predictive value of clinical-endoscopic assessment, MRI should be utilized as a targeted resource, reserved for cases of clinical ambiguity.

In conclusion, otoendoscopy is a fundamental diagnostic cornerstone for "real-life" cholesteatoma follow-up due to its precision, reproducibility, and immediate availability. Advanced radiologic investigations find their greatest utility within an integrated and multidisciplinary approach, rather than as a routine replacement for clinical expertise. Despite the inherent limitations of a retrospective study, these results outline a pragmatic diagnostic model that prioritizes efficient outpatient tools, ultimately optimizing patient care while reducing the long-term economic burden on the healthcare system.

REFERENCES

1. Adriaansens C, Bekkers S, Aarts MCJ. Determinants influencing cholesteatoma recurrence in daily practice: a retrospective analysis. *J Laryngol Otol.* 2022;136(2):119-124. doi:10.1017/S0022215121003546
2. Møller PR, Pedersen CN, Grosfjeld LR, Faber CE, Djurhuus BD. Recurrence of Cholesteatoma - A Retrospective Study Including 1,006 Patients for More than 33 Years. *Int Arch Otorhinolaryngol.* 2020;24(01):e18-e23. doi:10.1055/s-0039-1697989
3. Iino Y, Imamura Y, Kojima C, Takegoshi S, Suzuki JI. Risk factors for recurrent and residual cholesteatoma in children determined by second stage operation. *Int J Pediatr Otorhinolaryngol.* 1998;46(1-2):57-65. doi:10.1016/S0165-5876(98)00126-8
4. Kerckhoffs KGP, Kommer MJB, van Strien THL, et al. The disease recurrence rate after the canal wall up or canal wall down technique in adults. *Laryngoscope.* 2016;126(4):980-987. doi:10.1002/lary.25591
5. Ho SY, Kveton JF. *Efficacy of the 2-Staged Procedure in the Management of Cholesteatoma.* Vol 129.; 2003.
6. Bu DD, Schwam ZG, Wanna GB, Perez E, Cosetti MK. Cost-Effectiveness of Diffusion-Weighted Magnetic Resonance Imaging Versus Second-Look Surgery in Treating Cholesteatoma: A Modeling Study. *Otology and Neurotology.* 2024;45(2):163-168. doi:10.1097/MAO.0000000000004076
7. Baba A, Matsushima S, Fukuda T, et al. Improved assessment of middle ear recurrent/residual cholesteatomas using temporal subtraction CT. *Jpn J Radiol.* 2022;40(3):271-278. doi:10.1007/s11604-021-01209-2
8. De Foer B, Vercruysse JP, Bernaerts A, et al. Detection of Postoperative Residual Cholesteatoma With Non-Echo-Planar Diffusion-Weighted Magnetic Resonance Imaging. *Otology & Neurotology.* 2008;29(4):513-517. doi:10.1097/MAO.0b013e31816c7c3b
9. Velthuis S, Van Everdingen KJ, Quak JJ, Colnot DR. The value of non echo planar, diffusion-weighted magnetic resonance imaging for the detection of residual or recurrent middle-ear cholesteatoma. *Journal of Laryngology and Otology.* 2014;128(7):599-603. doi:10.1017/S0022215114001418
10. Tierney PA, Pracy P, Blaney SPA, Bowdler DA. An assessment of the value of the preoperative computed tomography scans prior to otoendoscopic "second look" in intact canal wall mastoid surgery. *Clin Otolaryngol Allied Sci.* 1999;24(4):274-276. doi:10.1046/j.1365-2273.1999.00238.x
11. Pietraszek M, Stański M, Marszał J, et al. Detection of cholesteatoma recurrence by magnetic resonance imaging (DWI non-EPI sequence) – how can we minimise false results? *Pol J Radiol.* 2025;90:318-323. doi:10.5114/pjr/203991
12. Piekarek A, Zatoński T, Kolator M, Bladowska J, Sasiadek M, Zimny A. The value of different diffusion-weighted magnetic resonance techniques in the diagnosis of middle ear cholesteatoma. Is there still an indication for echo-planar diffusion-weighted imaging? *Pol J Radiol.* 2022;87:51-57. doi:10.5114/pjr.2022.113194
13. Aarts MCJ, Rovers MM, van der Veen EL, Schilder AGM, van der Heijden GJM, Grolman W. The diagnostic value of diffusion-weighted magnetic resonance imaging in detecting a residual cholesteatoma. *Otolaryngology-Head and Neck Surgery.* 2010;143(1):12-16. doi:10.1016/j.otohns.2010.03.023
14. Amoodi H, Mofti A, Fatani NH, Alhatem H, Zabidi A, Ibrahim M. Non-echo Planar Diffusion-Weighted Imaging in the Detection of Recurrent or Residual Cholesteatoma: A Systematic

Review and Meta-Analysis of Diagnostic Studies. *Cureus*. Published online December 2, 2022. doi:10.7759/cureus.32127

15. Jindal M, Riskalla A, Jiang D, Connor S, O'Connor AF. A systematic review of diffusion-weighted magnetic resonance imaging in the assessment of postoperative cholesteatoma. *Otol Neurotol*. 2011;32(8):1243-1249. doi:10.1097/MAO.0B013E31822E938D
16. De Foer B, Vercruysse JP, Bernaerts A, et al. Middle ear cholesteatoma: non-echo-planar diffusion-weighted MR imaging versus delayed gadolinium-enhanced T1-weighted MR imaging-value in detection. *Radiology*. 2010;255(3):866-872. doi:10.1148/RADIOL.10091140
17. Silvola J, Palva T. Long-term results of pediatric primary one-stage cholesteatoma surgery. *Int J Pediatr Otorhinolaryngol*. 1999;48(2):101-107. doi:10.1016/S0165-5876(98)00188-8
18. Stangerup SE, Drozdiewicz D, Tos M. Cholesteatoma in children, predictors and calculation of recurrence rates. *Int J Pediatr Otorhinolaryngol*. 1999;49 Suppl 1(SUPPL. 1). doi:10.1016/S0165-5876(99)00136-6