Research Article

Incidental Thyroid Carcinoma in Patients Undergoing Surgery for Presumed Benign Multinodular Goitre: A Systematic Review and Meta-Analysis

Mousa Basim Mousa Alqadan[1], Reef Haroun Ahmad Al-Qaisi[2], Amro Ra'ed AbdAljaleel AlMbidin[3], Abdallah Daradkeh[4], Rama J. Alawneh[5], Abeer A. Alsarairah[6]

[1,5]Department of General Surgery, Faculty of Medicine, Mutah University, Al-Karak, Jordan
[6]Department of Internal Medicine, Faculty of Medicine, Mutah University, Al-Karak, Jordan
***Corresponding Author**: Mousa Basim Mousa Alqadan, Mail id: Mosaalqadan1234@gmail.com

Abstract

Background: Incidental thyroid carcinoma is sometimes identified after surgery for presumed benign multinodular goitre (MNG). Reported detection proportions vary widely, limiting their use in preoperative counselling. This review estimated the postoperative carcinoma proportion and examined its histology and sources of variation. **Methods:** PubMed, Scopus and Web of Science were searched for peer-reviewed studies published from 12 November 2011 to 12 November 2026. Two independent authors assessed full texts and extracted eligible study data. Patient-level carcinoma counts from MNG-specific surgical cohorts were synthesized using a random-effects meta-analysis of logit-transformed proportions. Heterogeneity, restricted-subgroup estimates, leave-one-out influence and exploratory clinicopathological associations were examined. **Results:** Of 2,233 retrieved records, 1,454 duplicates and 622 title/abstract records were removed. Of 157 full-text reports reviewed, 146 were excluded because eligibility criteria were unmet and/or extractable quantitative data were insufficient. Eleven cohorts (7,400 surgical patients; 1,449 incidental carcinomas) were included. The pooled postoperative carcinoma proportion was 13.7% (95% confidence interval [CI], 7.9%–22.7%), with substantial heterogeneity (I^2 = 98.2%). Excluding toxic-MNG-only cohorts yielded 12.8% (95% CI, 6.8%–22.8%). Papillary carcinoma predominated where histology was reported. In two studies, the pooled crude odds ratio for male versus female sex was 1.27 (95% CI, 1.04–1.56); other risk-factor results were largely study-specific. **Conclusions:** Incidental carcinoma is a measurable postoperative finding in selected patients undergoing surgery for presumed benign MNG, but differences among cohorts are too large for a single percentage to predict individual risk. Histological subtype, preoperative selection and surgical context should guide interpretation. The findings do not support routine extensive surgery solely on the basis of the pooled proportion.

Keywords: Multinodular goitre; incidental thyroid carcinoma; thyroidectomy; papillary thyroid microcarcinoma; systematic review; meta-analysis

1. Introduction

Multinodular goitre (MNG) is a common form of thyroid nodular disease. The majority of thyroid nodules are non-cancerous and many do not need to be treated. Surgical treatment is indicated if the goitre is causing pressure symptoms, when it is still growing, when there is a problem with its functioning or if malignancy is suspected. The typical evaluation for patient management prior to surgery typically involves clinical assessment, ultrasound, thyroid function tests, and selected fine needle aspirates (FNA) (Durante et al., 2023; Haugen et al., 2016).

Even in apparently benign nodular disease thyroid carcinoma is an important issue that must be addressed. Thyroid cancer is a very large burden according to global cancer surveillance but the rate of cancer detected is not a direct indicator of thyroid cancer with clinical implications (Bray et al., 2024). With increasing use of ultrasound and other imaging modalities, the number of small papillary cancers that are being found has also increased – a lot of these are indolent. The significance of the interpretation of a cancer encountered serendipitously following thyroid surgery is particularly important, given the risk of overdiagnosis (Durante et al., 2023; Chen & Haymart, 2026).

In the current review, incidental thyroid carcinoma means that the thyroid cancer was diagnosed only on final pathological examination following surgery which was performed for presumed benign MNG. A detection outcome that occurs following surgery which is in a selected surgical population. It should not be mistaken with population prevalence of thyroid cancer, or with the prevalence of thyroid cancer in all people with thyroid nodule(s) (Apostolou et al., 2021; Kaliszewski et al., 2016).

If a patient has a benign pretreatment determination, not all of the nodules in a multinodular gland are benign. Although ultrasound can be useful for determining those that need to be sampled, it is hard to obtain a complete pathological sampling of several nodules prior to surgery. FNA is used to look at certain areas of the thyroid and not the entire thyroid. Small malignant foci may thus go undetected on the pre-op biopsy, and only be identified when the surgical specimen is examined (Campbell et al., 2015; Durante et al., 2023; Haugen et al., 2016).

The postoperative rate of cancer has been reported in published surgical series and the rate is quite variable. In 2016, Kaliszewski et al. found incidental carcinoma in 2.1% (49 out of 2,306) of patients who had surgery for presumed benign MNG. In comparison, according to Apostolou et al. (2021), 31.7% of 3233 patients who underwent surgery developed

cancer (1026 cases). These variations could be attributed to referral patterns, the definition of benign disease, thyroid function status, sampling of the thyroid gland before surgery and ascertainment of the histopathology. These should not be thought of as just shifts in the underlying population risk (Apostolou et al., 2021; Kaliszewski et al., 2016; Pavone et al., 2026).

The kind of unexpected carcinoma also is important. In postoperative series of MNG, papillary thyroid carcinoma is a common finding, many times in the form of papillary thyroid microcarcinoma. When microcarcinoma is discovered, it may not be the indication of surgery but rather the pathological finding, and it is not known that a more extensive surgery would have resulted in better outcomes. Hence, it is important to handle histological definitions and differential diagnosis between affected patient and multiple tumour foci with care (Arican et al., 2023; Chen & Haymart, 2026; Fama et al., 2018).

While a reliable synthesis is important, it needs to take into account both aspects of the surgical decision. Extensive thyroidectomy could decrease the risk of recurrence of goitre or the need for completion thyroidectomy in certain individuals and could increase the transient postoperative complications. The proportion of postoperative malignancy in a pool should NOT be the ONLY indication for total thyroidectomy, but should be used to inform counselling and risk assessment (Al-Hakami et al., 2024; Bharath et al., 2023; Kaliszewski et al., 2016).

The overall goal of this systematic review and meta-analysis was to estimate the prevalence of incidental thyroid carcinoma (ITC) found in patients who have undergone surgery for presumed benign MNG in the last 15 years. Secondary objectives included to describe the reported histological patterns, investigate between-study heterogeneity, to test the robustness of the pooled estimate to clinical and methodological restrictions, and to explore risk-factor associations when appropriate numbers of counts were available. We hypothesized that unexpectedly identified carcinoma would be a consistent finding in these surgical cohorts, that papillary carcinoma would be the most common type, and that the proportions of the various types would differ significantly between studies, due to the differences in the study selection and the preoperative evaluation (Apostolou et al., 2021; Campbell et al., 2015; Chen & Chen, 2022; Kaliszewski et al., 2016).

2. Literature Review

The first step of the preoperative evaluation of MNG is to determine which nodule(s) require further evaluation and if the patient requires treatment. The current guidelines focus on risk oriented ultrasound evaluation and risk oriented FNA and the correlation of FNA and clinical data. When the nodule is low risk, a benign cytology result is usually a good indicator for conservative management. Even with benign goitre, surgery may be indicated due to compression, growth of the goitre, thyroid function or patient preference. Therefore, surgical series after surgery will not contain all individuals with MNG, since they are a selected group of patients who meet clinical criteria.

Several nodules may coexist and diagnostic limitations are of particular interest. There is no guarantee that sampling the largest (or most suspect) nodule rules out the presence of a malignant focus in another part of the gland. Campbell et al. (2015) investigated 134 patients who had thyroid surgery for non-functioning cervical MNG in spite of a benign FNA previous to surgery. Cancer was diagnosed in 31 patients (23.1%) with 21 having micro-papillary cancer. Moreover, the authors found an association between male sex and non-micropapillary carcinoma, as well as between the source of cytology review and non-micropapillary carcinoma. The findings, however, do not reflect that benign FNA in the larger thyroid-nodule population has a 23% false-negative rate (Campbell et al., 2015; Haugen et al., 2016).

Some studies may involve a benign cytology whereas others may involve a pre-operative impression of benign MNG. In South Africa, Bombil et al. (2014) found that 5.7% of 70 patients who had FNA and had been diagnosed with benign MNG on their thyroid surgery specimen had incidental papillary cancer. Kaliszewski et al., (2016) have found 49 unsuspected malignancies in surgical MNG patients in Poland, who had no clinical suspicion of malignancy. The reported proportions cannot be compared without taking into account the different sampling frames, inclusion rules and protocols for examining the tissues (Bombil et al., 2014, Kaliszewski et al., 2016).

The percentages are much higher in other surgical series. Apostolou et al. (2021) analysed a big tertiary centre database of patients who underwent thyroid surgery for benign MNG and they reported incidental thyroid cancer in 31.7% of the patients. They correlated malignancy with various demographic, functional and histological characteristics such as thyroiditis and smaller thyroid size. But these associations were developed among a selected surgical population, and may not necessarily be applied to patients who are not treated surgically. Part of the differences in the estimates exists between hospitals, which may be attributed to variation in surgeon referral rates and disease severity (Apostolou et al., 2021; Pavone et al., 2026).

There has been a renewed focus on the type of malignancy; not the number of malignant specimens. Fama et al. (2018) reviewed papillary cancers that were not detected prior to thyroid surgery for MNG that were considered benign. Arican

et al (2023) separately mentioned both papillary carcinoma and papillary microcarcinoma and pointed out the fact that small tumours may represent a significant proportion of incidental tumours. In a similar study, Chen and Chen (2022) also studied clinicopathological parameters in euthyroid patients who had total thyroidectomy for MNG. These designs highlight the need for a review to differentiate all incidental carcinomas from papillary carcinoma and papillary microcarcinoma and present denominators for each endpoint (Arican et al., 2023; Chen & Chen, 2022; Fama et al., 2018).

Histology should not be assumed as the clinical significance of papillary microcarcinoma. There are some small papillary lesions that have an indolent course and modern thyroid-cancer management increasingly takes into account tumour characteristics, the benefits of a treatment and its harm. The discovery of a microscopic focus at the time of goitre surgery may change the subsequent management, but does not prove that the patient who is initially discovered to be asymptomatic would have benefited from surgery on the thyroid. This issue of overdiagnosis has therefore a direct impact on the interpretation of postoperative cancer rates (Chen & Haymart, 2026; Durante et al., 2023; Haugen et al., 2016).

Another source of the clinical variation is functional status. There are certain differences between toxic multi nodular goitre and euthyroid MNG with respect to their indications for surgery and the pre-operative work-up. Malignant findings were reported after surgery in the study on toxic nodular disease done by Mohamed et al. (2022) but the study also included solitary toxic nodules. Only groups which are separately extractable and multinodular can provide an answer to the present review question. If a mixed cohort with toxic nodules and toxicity were treated like all patients had MNG, this would lead to an inaccurate numerator and denominator (Durante et al., 2023; Mohamed et al., 2022).

A related, but different, research question relates to surgical extent. Kaliszewski et al. (2016) mentioned the options of completion surgery if cancer was encountered following less radical surgeries. However, comparative data on the benign MNG shows a trade-off between recurrence and morbidity of surgery. Bharath et al. (2023) performed a meta-analysis of randomized trials to compare total thyroidectomy with less than total thyroidectomy for benign non-toxic MNG. Other synthesis reported a reduction in recurrent events following total thyroidectomy compared to partial thyroidectomy, but an increased rate of transient adverse events following total thyroidectomy (Al-Hakami et al., 2024), with no significant difference in thyroid-cancer incidence. The estimate of incidental cancer should thus be viewed in the context of the surgical risks and advantages (Bharath et al., 2023; Al-Hakami et al., 2024).

There are significant methodological problems in this literature. There is variation among studies in the criteria for surgery, definition of tumour, the taking of samples of the histology and whether patients with suspicious cytology were excluded prior to analysis. Simple pooled numerators and denominators can be influenced by larger cohorts and small, but narrowly drawn cohorts, can give different rates. In the case of a random-effects synthesis, between-study variability is made explicit, but when the between-study variability is very large, the single summary proportion is of little value for clinical prediction. Moreover, methodological studies also warn that the traditional two-step approach to the meta-analysis of proportions is not without limitations, and that in the future, generalized linear mixed models are another option for robustness-checking (DerSimonian & Laird, 1986; Lin & Chu, 2020).

All of the published evidence taken together suggests that a targeted synthesis focusing on surgery for presumed benign MNG, and not all thyroidectomies or all nodular thyroid disorders is necessary. The denominator (patient-level) should be reported appropriately, MNG-only subgroups should be maintained from mixed cohorts, all-carcinoma end points should be distinguished from sub-type specific end points, and the impact of toxic disease and benign-FNA criteria should be explored. It should also not assume that observational associations represent causation and recognize that very selective surgical data (Apostolou et al., 2021; Bombil et al., 2014; Campbell et al., 2015; Lin & Chu, 2020) can be misleading.

3. Methodology

3.1 Study design and objective

The aim of this systematic review and meta-analysis was to assess the incidence of thyroid carcinoma in surgeries performed for presumed benign MNG (multinodular goitre). The main endpoints were the percentage of the eligible surgical patients in whom thyroid carcinoma was established on the postoperative pathological examination without the preoperative diagnosis of malignancy. The review was conducted using the PRISMA 2020 reporting guidance (Page et al., 2021). Event counts at the study level were used for analyses, and individual-patient data were not available.

3.2 Search strategy and information sources

Searching was performed in original peer-reviewed publications from November 12, 2011 to November 12, 2026 in PubMed, Scopus and Web of Science-Core Collection. Database-specific strategies took into account multinodular or nodular goitre (including both spellings “goiter”) and malignancy and postoperative, histopathological, incidental, occult

or unexpected detection. Search expressions are given in the Appendix. The number of records found on PubMed was 448, from Scopus was 1198 and from Web of Science was 587. The targeted terminology might not have been as sensitive to the incidental diagnosis of cancer in studies with only full-text available.

3.3 Eligibility criteria and data sources

The studies selected for inclusion had to meet the following criteria: original surgical observational cohorts with a definable population undergoing thyroid surgery with a presumed benign indication of MNG, from which an extractable number of patients with incidental thyroid carcinoma could be obtained. For mixed nodule or mixed-thyroidectomy reports the separately extractable groups of MNG were retained. Two toxic-MNG-only cohorts were added to the main dataset as the data for MNGs was available and were excluded in a sensitivity analysis. Those without separable counts, duplicates and reports without a numerator and denominator were excluded and those with known malignant thyroid disease were studied.

The extraction fields for information from the study levels were: publication year, country, surgical cohort characteristics, pre-operative benign assessment, surgical procedures, number of patients eligible for denominator, number of incidental cases of carcinoma, histological types and 2×2 counts reported by sex or thyroid function. Patients with carcinoma were counted as such (subtype counts were not reported and counted as missing).

3.4 Selection process and PRISMA flow

There were 2,233 records found in the searches. 779 titles and abstracts were left to be screened after the removal of 1,454 duplicates, of which 622 were excluded. Two independent authors performed full-text eligibility assessment and data-extraction assessment of the remaining 157 reports. Overall, 146 reports were excluded due to not meeting the eligibility criteria and/or not having extractable quantitative data. One primary meta-analysis was conducted which included 11 studies. The study-selection process is shown in Figure 1.

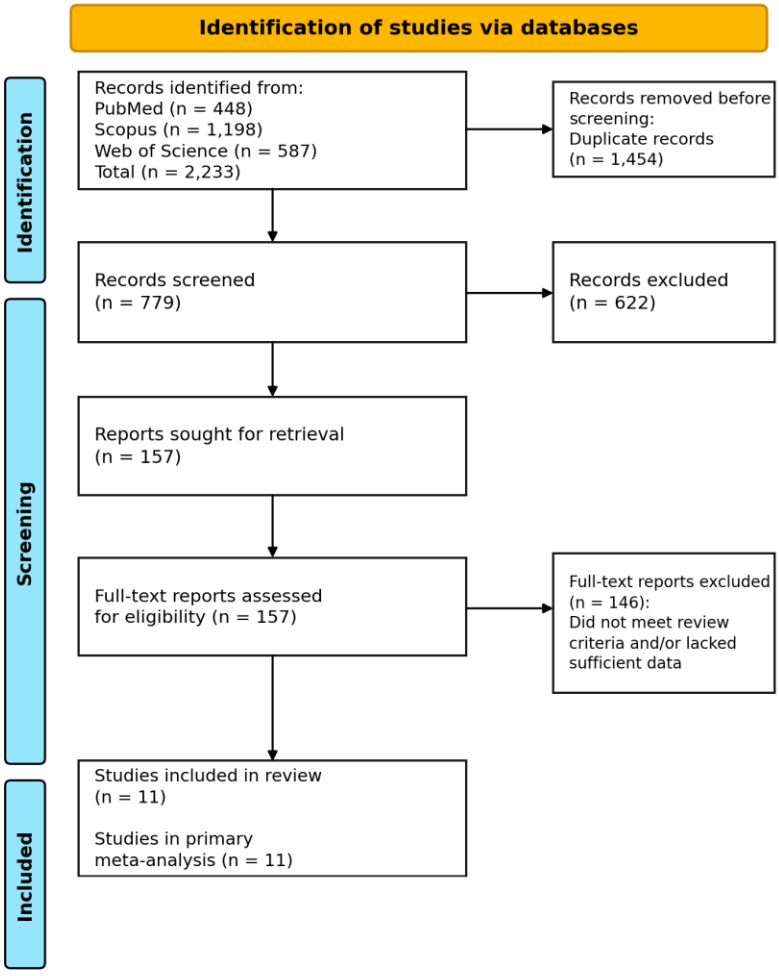


Figure: 1 PRISMA 2020 flow diagram of record identification, screening, eligibility assessment, and inclusion.

3.5 Statistical analysis

The outcome in each study was an incidence proportion x/n and was logit transformed as $y = \log[x/(n - x)]$ with within-study sampling variance $v = 1/x + 1/(n - x)$. The DerSimonian–Laird estimator was applied for between-study variance (τ^2) and inverse-variance weights, $1/(v + \tau^2)$, were applied. A pooled logit and normal approximation 95% CI (confidence interval) were computed and back transformed. In the forest plots, 95% CIs for the study level were obtained as Wilson score intervals (as implemented in R). This two-step approach was based on the traditional random-effects approach; however, there are limitations to pooling proportions (DerSimonian & Laird, 1986; Lin & Chu, 2020).

Cochran's Q test, I^2 and $\tau 2$ were used to study heterogeneity. The random-effects model was used to provide a descriptive 95% prediction interval on the logit scale. Restricted maximum likelihood (REML) estimation was compared to the primary DerSimonian–Laird approach, Hartung–Knapp limits, leave-one-study-out estimates, exclusion of toxic-MNG-only populations and restrictions on publication period and cohort size in sensitivity analyses. These analyses were exploratory without the intention of causal inference between sub-groups. Only studies that clearly extracted patient-level events for each outcome were included in the analyses of secondary papillary thyroid microcarcinoma (PTMC) and papillary thyroid carcinoma (PTC).

Crude odds ratios (ORs) and risk ratios (RRs) with 95% CIs were calculated from 2×2 count data reported. More than one study included comparisons of males versus females; these were pooled in a separate exploratory random-effects analysis. Other contrasts of risk factors were reported separately due to the different populations and exposures. A funnel plot was descriptively examined, but not formally tested for evidence of publication bias given the large amount of heterogeneity (Sterne et al., 2011). The statistical analyses were performed in Python with NumPy, SciPy and the graphical part was created with Matplotlib.

4. Results

4.1 Identification of eligible studies

The final searches yielded 2,233 records (PubMed, 448; Scopus, 1,198; Web of Science, 587). 779 records were title/abstract screened and 622 were excluded after 1,454 duplicates have been removed. The 157 remaining full-text reports were then independently analysed by two authors who extracted data. A total of 146 studies were excluded due to ineligibility and/or insufficient quantitative data that could be extracted, resulting in 11 studies for quantitative synthesis (Figure 1).

4.2 Study characteristics

The 11 cohorts (2013–2026) were from 10 countries. The denominators for cohorts varied between 61 and 3,233 patients. A total of 7,400 surgical patients and 1,449 patients with incidental carcinoma were included in the combined cohort, which gave a simple patient weighted rate of 19.6%. This crudely estimated ratio is different from the random-effects pooled estimate because there were big differences in the sizes and reported incidences of cohorts. Table 1 shows summary information regarding the characteristics and dominant counts.

Table: 1 Characteristics and observed proportions of the included surgical cohorts.

Study	Country	Eligible N	Cases, n	Rate, %	Preoperative population
Gál et al. (2013)	Hungary	264	9	3.4	Presumed benign MNG
Bombil et al. (2014)	South Africa	70	4	5.7	Benign FNA
Campbell et al. (2015)	United States	134	31	23.1	Benign FNA
Kaliszewski et al. (2016)	Poland	2,306	49	2.1	Presumed benign MNG
Fama et al. (2018)	France	207	25	12.1	Presumed benign MNG
Özpek & Topal (2019)	Turkey	114	13	11.4	Toxic MNG
Apostolou et al. (2021)	Serbia	3,233	1026	31.7	Presumed benign MNG
Mohamed et al. (2022)	Egypt	61	17	27.9	Toxic MNG
Chen & Chen (2022)	Taiwan	72	19	26.4	Benign FNA
Arican et al. (2023)	Turkey	716	224	31.3	Presumed benign MNG
Pavone et al. (2026)	Italy	223	32	14.3	Presumed benign MNG

All counts in Table 1 refer to the number of eligible MNG surgical patients and are not the total number of mixed-thyroidectomy patients as the original study included additional indications. The toxic-MNG group was isolated from the other nodules in the study of Özpek and Topal (2019), for example. The between-study variation in the benign cytology

criteria, thyroid-function status and type of surgery is large, as discussed in the main forest plot and robustness analyses below.

4.3 Primary prevalence meta-analysis

An incidental carcinoma proportion of 13.7% (95% CI [7.9%, 22.7%]) was estimated in the 11-study random-effects meta-analysis (Figure 2). The estimate was lower than the crude patient-weighted proportion of 19.6% due to weighting of individual cohorts, and the large between-study variance. In Fig. 2 the surface area of the squares is proportional to the random-effects inverse-variance weights; horizontal lines represent the 95% Wilson CIs for the individual studies, and the diamond represents the pooled estimate.

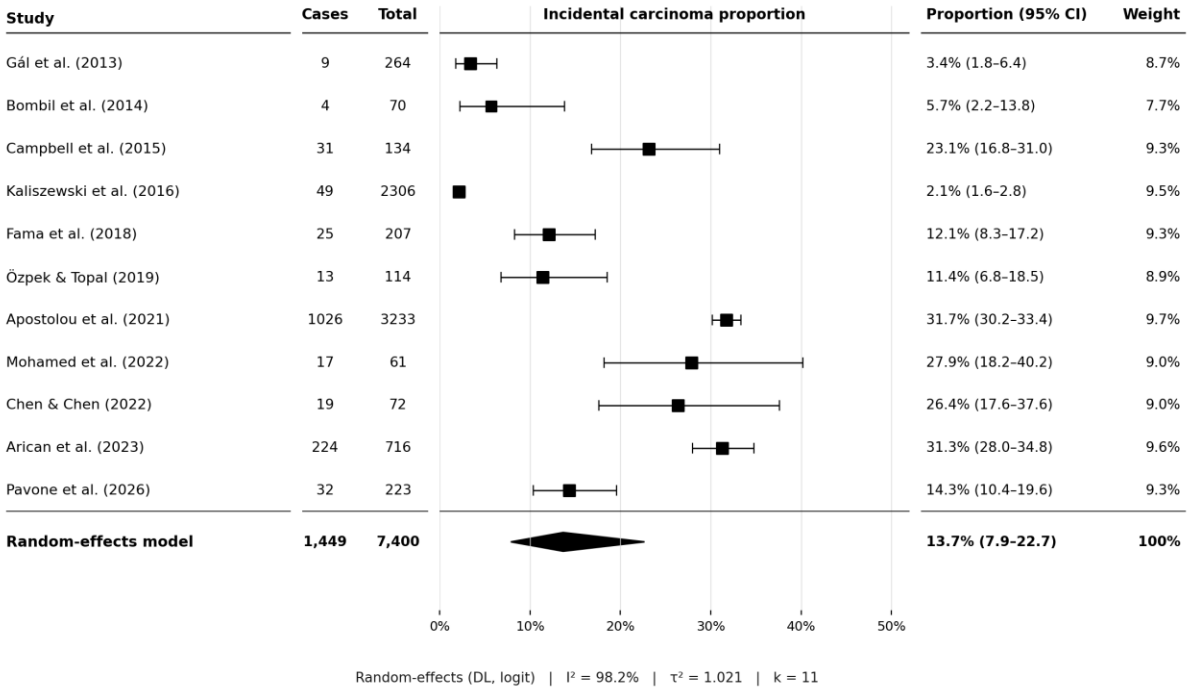


Figure: 2 Random-effects forest plot of incidental thyroid carcinoma proportions after MNG surgery. Squares represent estimates from the studies, bars represent Wilson 95% confidence intervals, and the diamond represents the pooled DerSimonian–Laird estimate.

Heterogeneity was pronounced (Q = 540.80, df = 10, p < .001; I^2 = 98.2%; τ^2 = 1.021, logit scale). There was a wide estimated 95% prediction interval of 1.4% to 63.4% suggesting that rates in a new comparable setting may vary widely from the summary mean. These are approximate, and should be interpreted with caution due to methodological differences between surgical cohorts.

4.4 Sensitivity and influence analyses

REML estimation gave a pooled proportion of 13.7% (95% CI [7.7%, 23.0%]) (τ^2 = 1.084). The primary model's confidence limits went from 7.2% to 24.5% after Hartung–Knapp uncertainty adjustment. The summary of the exploratory restriction on the eligibility of the study and the time period is shown in Table 2. These are robustness checks and not independent comparisons of equivalently defined clinical populations.

Table: 2 Random-effects sensitivity analyses across clinically defined study subsets.

Population subset	k	Total N	Cases	Pooled % (95% CI)	I^2, %
All included MNG cohorts	11	7,400	1449	13.7 (7.9–22.7)	98.2
Excluding toxic-MNG-only cohorts	9	7,225	1419	12.8 (6.8–22.8)	98.5
Benign-FNA-defined cohorts	4	390	67	15.9 (9.0–26.5)	80.1
Published 2021–2026	5	4,305	1318	26.4 (21.3–32.2)	85.9
Cohort size ≥ 200	6	6,949	1365	11.3 (4.9–23.7)	99.0

The benign-FNA-defined subset included the separable populations reported by Bombil et al. (2014), Campbell et al. (2015), Özpek and Topal (2019), and Chen and Chen (2022). The estimates for the five subset definitions are shown in Figure 3; changes in pooled incidence largely reflect clinical selection and cohort composition and cannot establish the direction of bias in any particular study.

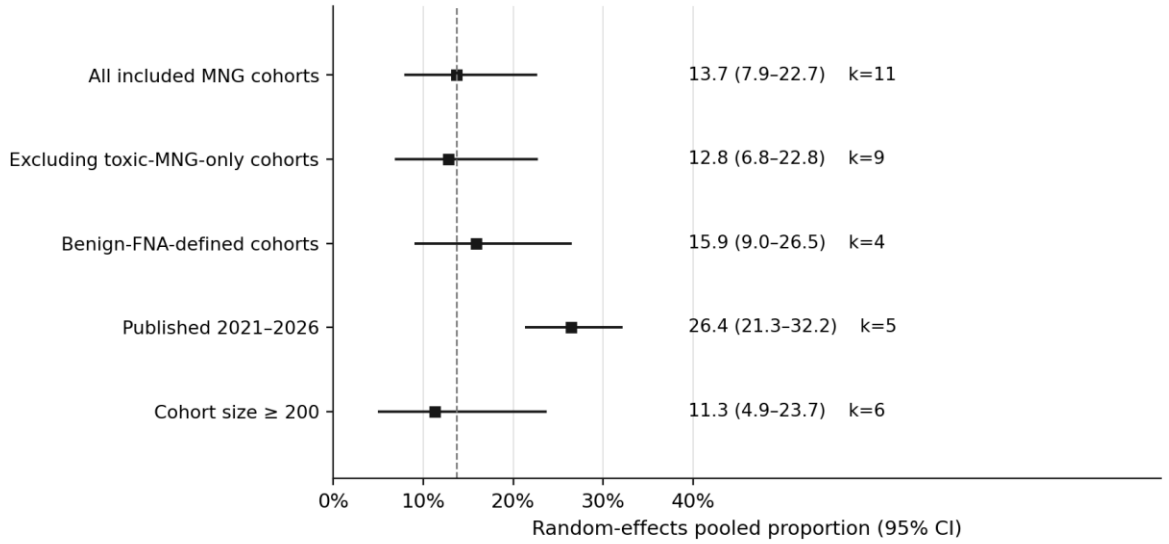


Figure: 3 Forest-style comparison of the random-effects pooled estimates across sensitivity subsets. The vertical reference line denotes the overall primary estimate.

Leave-one-out pooling produced estimates ranging from 12.3% to 17.1%. Excluding Apostolou et al. (2021) yielded the lowest pooled estimate, whereas excluding Kaliszewski et al. (2016) produced the highest. Figure 4 displays the re-estimated pooled proportion after omission of each study, indicating the extent to which the headline finding depends on particular cohorts.

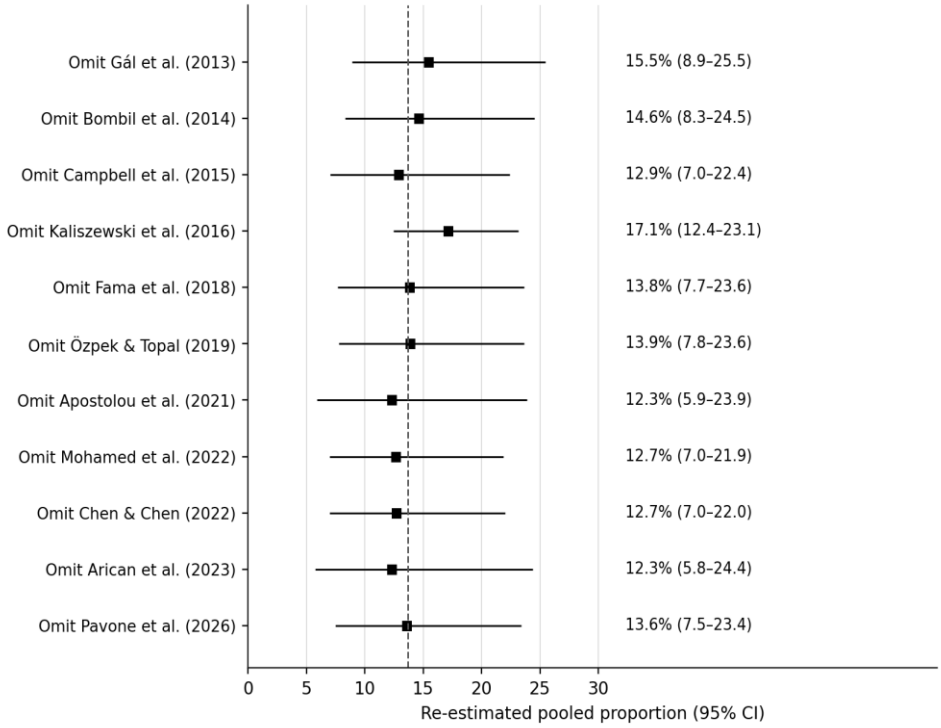


Figure: 4 Leave-one-out influence analysis Each row represents the random-effects pooled incidence proportion after removing the named study; the dashed vertical line represents the full-sample pooled estimate.

Although the point estimate changed across restrictions and omissions, high heterogeneity persisted, and the available study-level covariates were insufficient for a credible adjusted meta-regression. Accordingly, subgroup patterns are interpreted descriptively.

4.5 Histopathological outcomes

Separate patient level data for papillary thyroid carcinoma (including microcarcinoma) was available in eight cohorts and explicit counts of PTMC patients in four cohorts. The exploratory pooled papillary carcinoma proportion among studies reporting this outcome was 12.0% (95% CI [6.3%, 21.7%]) (I^2 = 98.4%); the corresponding PTMC estimate was 17.8% (95% CI [16.7%, 19.0%]) (I^2 = 0.0%). These two estimates are derived from different subpopulations of reporters and PTMC is not a distinct disease category. The study level outcome estimates without imputation of missing subtype data are shown in Figure 5.

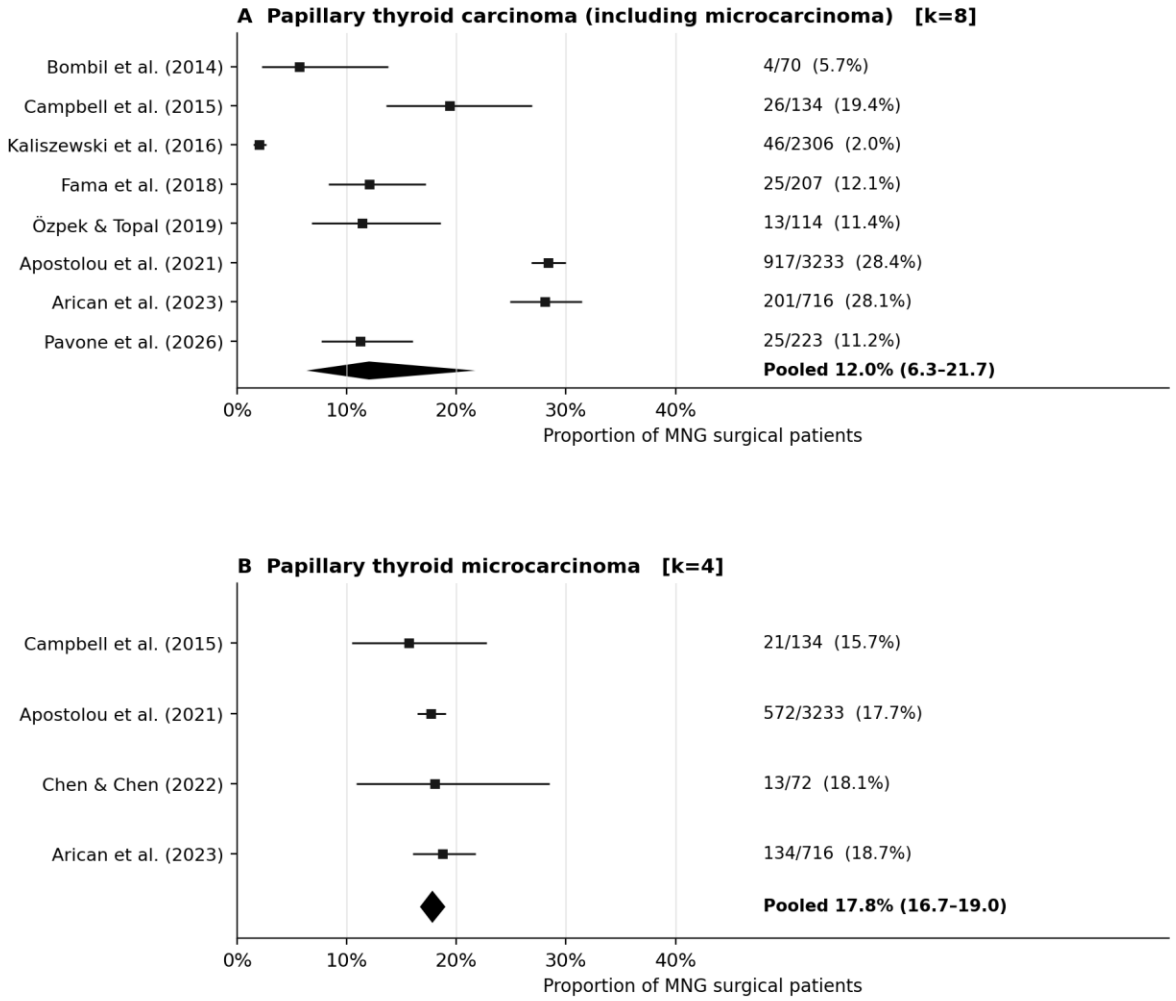


Figure: 5 Histology-specific forest plots for (A) papillary thyroid carcinoma and (B) papillary thyroid microcarcinoma. The included study sets differ because of reporting completeness.

Some other malignant variants reported were follicular, oncocytic/Hürthle-cell, medullary or anaplastic carcinomas, but the definitions and detection of these variants differed in the original reports. Arican et al. (2023) noted 18 follicular, three oncocytic and two medullary cancers in a total of 716 eligible patients, while Apostolou et al. (2021) reported 17 follicular, 65 oncocytic/Hürthle-cell, 21 medullary and six anaplastic cancers in 3,233 patients. Because of variable reporting, these lesser common categories were not pooled. Patients with multifocal tumour deposits were not considered as multiple.

4.6 Exploratory risk-factor associations

There were two reports containing extractable numbers of male and female carcinomas (3,456 participants). The pooled unadjusted OR for male versus female sex was 1.27 (95% CI [1.04, 1.56]; I^2 = 0.0%). Inference is limited by the

confidence interval and only two studies are available. The crude ORs and RRs have been calculated directly for all six possible 2×2 contrasts without combining non-comparable exposures as shown in table 3.

Table: 3 Crude risk-factor comparisons from available within-study two-by-two cells.

Contrast and study	Exposed cases/N	Reference cases/N	OR (95% CI)	RR (95% CI)
Male vs female; Apostolou et al. (2021)	161/445	865/2788	1.26 (1.02–1.55)	1.17 (1.02–1.33)
Male vs female; Pavone et al. (2026)	8/42	24/181	1.54 (0.64–3.72)	1.44 (0.69–2.97)
Hashimoto thyroiditis vs none; Apostolou et al. (2021)	183/385	843/2848	2.15 (1.74–2.67)	1.61 (1.43–1.81)
Chronic nonspecific thyroiditis vs none; Apostolou et al. (2021)	138/284	888/2949	2.19 (1.72–2.80)	1.61 (1.41–1.84)
Euthyroid vs hyperthyroid; Apostolou et al. (2021)	814/2360	175/784	1.83 (1.52–2.21)	1.55 (1.34–1.78)
Hypothyroid vs euthyroid; Apostolou et al. (2021)	37/89	814/2360	1.35 (0.88–2.08)	1.21 (0.94–1.55)

Only the male-versus-female contrast is repeated across two studies. All other comparisons arise from Apostolou et al. (2021), and cannot be pooled as independent study-level estimates. The odds ratios for Hashimoto thyroiditis, chronic nonspecific thyroiditis and thyroid-function contrasts are displayed on a logarithmic scale in Figure 6. These associations are unadjusted and may be confounded by preoperative referral decisions and underlying pathology.

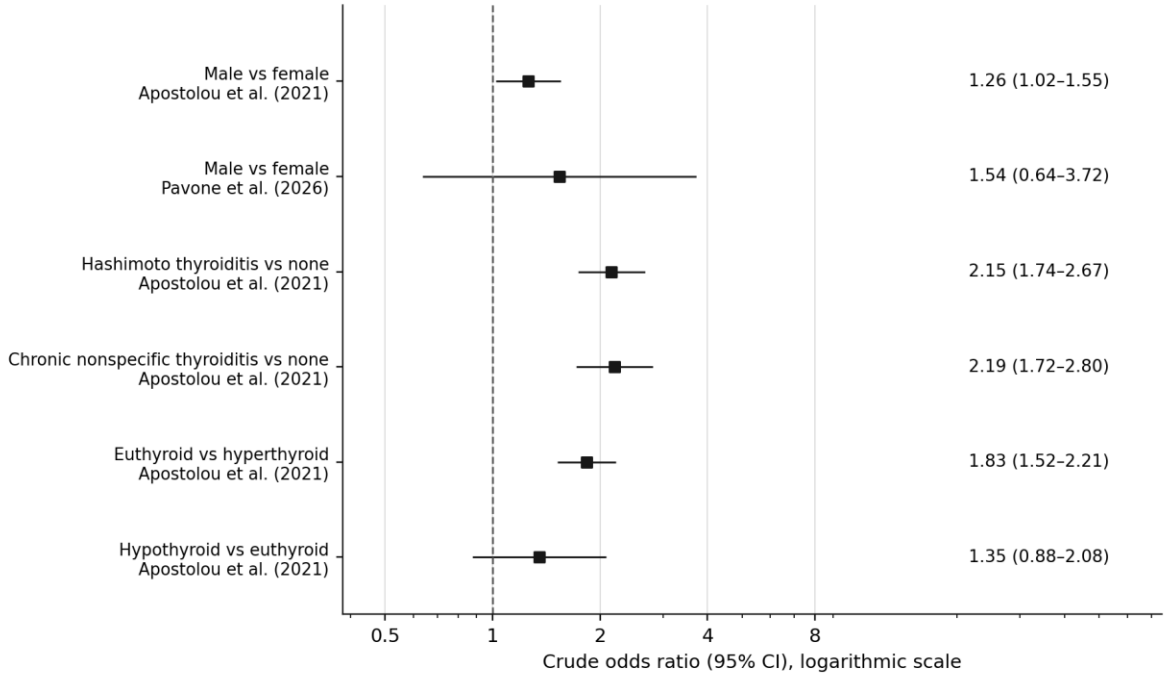


Figure: 6 Forest plot of crude odds ratios and 95% confidence intervals. The vertical line at OR = 1 indicates no association; rows represent distinct, non-equivalent clinical contrasts.

The number of studies with extractable covariate data was too small for meaningful multivariable adjustment, evaluation of interactions or risk-factor meta-regression. Consequently, these results are presented as hypothesis-generating evidence only.

4.7 Exploratory assessment of small-study effects

A funnel plot of logit-transformed study-specific estimates against their standard errors is provided in Figure 7. The between-study heterogeneity was extreme (I^2 above 98%), and the dependence between estimated prevalence and

standard error creates additional difficulties for interpreting this type of plot. As recommended in methodological guidance, asymmetry cannot be attributed to selective publication without considering other sources of heterogeneity (Sterne et al., 2011).

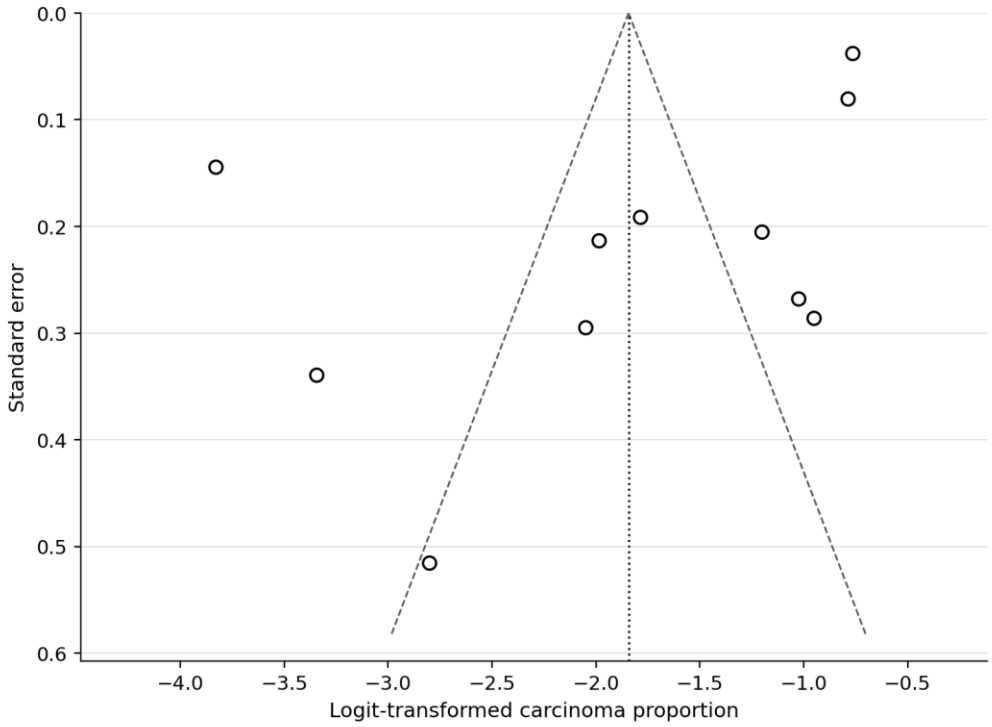


Figure: 7 Exploratory funnel plot of the primary cohort proportions on the logit scale. Dashed lines show nominal 95% reference limits around the pooled estimate; they do not constitute a formal test of publication bias.

Formal tests of funnel plot asymmetry were not performed because the small number of studies and extreme heterogeneity limit their interpretability.

4.8 Synthesis of main findings

Across 11 observational surgical MNG cohorts including 7,400 patients, there were 1,449 incidentally identified carcinomas. The pooled random-effects estimate was 13.7% (95% CI [7.9%, 22.7%]), accompanied by severe heterogeneity (I^2 = 98.2%). Reported PTC and PTMC proportions were analyzable in smaller, non-identical subsets, and the evidence for clinical predictors was limited to a few unadjusted 2×2 comparisons. These findings quantify the variation in postoperative carcinoma detection under different preoperative benign definitions but should not be interpreted as the risk among all patients with MNG in the general population.

5. Discussion

This synthesis found the incidental thyroid carcinoma in 11 surgical MNG cohorts comprised of 7,400 patients. The random-effects pooled detection proportion was 13.7% (95% CI, 7.9%–22.7%). The study's main hypothesis is that carcinoma can go undiagnosed prior to surgery for “benign” MNG, which is corroborated by this finding. There is however, significant heterogeneity (I^2 = 98.2%) and a large model-based prediction interval (1.4% – 63.4%) with the main finding. The average thus represents the evidence that is included rather than a reliable chance for an individual patient or a representative hospital. Another difference is that the crude proportion of 1,449/7,400 (19.6%) assigns more weight to the larger cohorts while the random-effects estimate is based on between-study variability. Both estimates are not a reflection on the level of thyroid cancer in the general population with MNG (Apostolou et al., 2021; Kaliszewski et al., 2016; Lin & Chu, 2020).

This difference between series of individual surgical procedures explains why it is important to use a pooled number with caution. Kaliszewski et al. (2016) demonstrated incidental malignancy in 49 of 2,306 patients (2.1%) who were picked up for surgery, although the malignancy was not suspected prior to surgery. Apostolou et al. (2021) found a total of 1026 cancers in 3233 patients (31.7%) at a tertiary endocrine surgery unit, excluding those with Bethesda III – VI cytology. This is too great a difference to be due to chance sampling. Case-mix, referral patterns, definitions of a benign

preoperative assessment, thyroid-function and operative indication and histological examination may all play a part. It was not possible in the present study to separate the effects of each of the factors and the different rates should not be interpreted as evidence for a better or worse diagnostic practice at one centre than the other.

The differences are still evident even in those groups that have a benign cytological result. Bombil et al., (2014) found four cancers in 70 eligible patients (5.7%) while Campbell et al., (2015) found 31 cancers in 134 patients (23.1%) despite the benign FNA. The proportion of the four benign-FNA defined cohorts in the current review was 15.9% (95% CI, 9.0%–26.5%) with a high degree of residual heterogeneity (I^2 = 80.1%). This cannot be termed as the false-negative rate of thyroid FNA. It's the percentage of carcinoma in a subset of patients who went on to thyroid surgery for other indications, and sometimes the malignancy was in nodules in which they were not sampled. Consistent with the management of other types of nodule, current guidance does not recommend that ultrasound FNA biopsy be used with a solitary test result, but rather that imaging, cytology and clinical indications should be integrated for selecting nodules to undergo FNA biopsy (Campbell et al., 2015; Durante et al., 2023).

The sensitivity analyses confirm the need to precisely define the population. After excluding the two toxic-MNG-only cohorts, the pooled estimate was only slightly affected, being 12.8% (95% CI, 6.8%–22.8%) with no change in heterogeneity. These five studies reported during 2021-2026 yielded a higher overall summary estimate of 26.4% (95% CI, 21.3% - 32.2%). However, the identity and size of the studies included and the clinical selection of both studies are a confounding factor. It does not indicate a growing trend of malignancy with time. Leaving one out, the estimates range from 12.3% to 17.1% and they are sensitive to individual cohorts but not the broader issue of varying study populations. These findings suggest that it is important to report clinically defined subgroups findings alongside (and not subservient to) the overall pooled proportion (Apostolou et al., 2021; Mohamed et al., 2022; Pavone et al., 2026).

The histopathology interpretation is essential as a second layer of interpretation. The papillary type was the most commonly reported type, which was also found in the distributions reported by Apostolou et al. (2021), Arican et al. (2023) and Kaliszewski et al. (2016). The pooled proportions for papillary thyroid carcinoma (eight studies) and papillary microcarcinoma (four studies) were 12.0% and 17.8% in the outcome subsets available, respectively. The difference is that these are NOT competing estimates as the studies and the reporting denominators are different and microcarcinoma is a subcategory of papillary carcinoma. This would not imply a biological contradiction in the apparently greater proportion of microcarcinoma, however. Arican et al. (2023) correlated the presence of microcarcinomas and papillary cancers (microcarcinoma 134, papillary cancer 67) in a detailed series of 716 operated patients. It is important to report microcarcinoma separately because, with this pathological diagnosis, it is not necessarily possible to predict aggressive tumour biology.

This distinction is relevant to how to interpret clinical burden of incidental disease. So, long-term observation is considered one of the options for appropriately selected, low-risk papillary thyroid microcarcinoma; however, there is no evidence that such a strategy may be applied to every papillary thyroid microcarcinoma found after goitre surgery (Ito & Miyauchi, 2023). Widely, rising thyroid cancer incidence has been driven by the detection of small, clinically unimportant (indolent) thyroid cancers that might not have given rise to symptoms or been diagnosed if not evaluated (Chen & Haymart, 2026). Tumour stage, aggressive variants, nodal involvement, recurrence, survival and subsequent treatment was not always reported in the present dataset. Thus, it cannot say what proportion of the 1,449 cancers it found would have been symptomatic or have a bad outcome without the surgery. This meta-analysis might overestimate what can be determined when all detected carcinomas are considered to be equivalent clinical events.

Another restraint is whether or not an incidental cancer estimate should affect the type of thyroidectomy performed. In the series of Kaliszewski et al. (2016) 21 of 49 patients with incidental cancer later underwent completion surgery, which provides an example of the potential consequences of limited surgery. However, this meta-analysis is not aimed at the comparisons of surgical technique, postoperative complications or long-term recurrence. They have shown that randomized-trial syntheses in benign multinodular goitre have demonstrated a lower rate of goitre recurrence following total thyroidectomy at the expense of some temporary adverse events, especially transient hypoparathyroidism (Temporary hypoparathyroidism (Al-Hakami et al., 2024; Bharath et al., 2023). Thus, it is not advisable to routinely perform a total thyroidectomy based on the postoperative cancer rate. The options and their choices are still influenced by the bilateral disease, compressive symptoms, thyroid function, imaging, operative risk as well as patient preference and specialist judgement (Durante et al., 2023).

There is limited risk-factor information available which provides only tentative guidance. Two cohorts provided sex-specific numbers of cases of carcinoma, giving a crude OR of 1.27 (95% CI 1.04-1.56) for men vs women. While the estimate indicates a moderate connection in those groups, it's based on just two studies and does not account for age, referral, imaging, thyroid function or other factors. The results of stronger unadjusted relationships with Hashimoto thyroiditis and with chronic nonspecific thyroiditis came from a single, large cohort, while it was not possible to combine

the results of the various contrasts of thyroid function meaningfully. Apostolou et al. (2021) reported multivariable association within their own centre and Campbell et al. (2015) reported an association with male sex in a more restricted surgical population and in the setting of non-micropapillary carcinoma. These observations are helpful in development of hypotheses but not in the development of a risk score or causal relationship. Especially, post-operative histological thyroiditis is not always a predictor, which can be used to help decide on surgery.

The current synthesis has several methodological advantages compared to a general review of the problem of malignancy in thyroid nodules. The principle denominator was limited to those patients who were operated on for presumed benign MNG and MNG-only groups were applied if other nodular conditions were included. Events were based on patients affected and not number of tumour foci. The complete-text reports were evaluated and data retrieved by two independent reviewers. Only counts explicitly reported in the secondary histology and risk-factor analyses were treated as such and no absences of counts were interpreted as a zero count. Furthermore, DerSimonian–Laird and restricted maximum likelihood variance estimation were compared and the uncertainty of Hartung–Knapp was studied, as well as the influence of individual studies. These options help to clarify the statistical uncertainty but are not able to overcome the issue of significant clinical heterogeneity (DerSimonian & Laird, 1986; Lin & Chu, 2020; Page et al., 2021).

Important limitations remain. The types of data used were all observational surgical series, most of which were from specialist centres. External validity is limited by selection in and to surgery and referral/verification biases may exist. There were no consistent definitions of presumed benign MNG, toxic status, preoperative FNA and histological classification. Only 11 cohorts were included in the meta-analysis and some subgroup and risk factor analyses were based on four or fewer studies. Exclusions were reported as full text and study-level risk of bias stratification was not performed. The design of a funnel plot would not be able to diagnose publication bias reliably in the presence of such a high level of heterogeneity nor would the two-stage logit model with DerSimonian–Laird variance estimation be able to provide correct results for heterogeneous proportions. Prediction interval is also only approximate. A binomial-normal generalized linear mixed model and standardized risk-of-bias assessment could further assess the strength of these results in the future (Lin & Chu, 2020; Sterne et al., 2011).

Future studies should prospectively define presumption of benign MNG, state if malignancy was ruled out by cytology or not, clarify status on toxicity and indicate the reason and type of surgery. All carcinoma, papillary carcinoma, microcarcinoma, aggressive features and clinically important outcomes should be distinguished in the consistent patient level core dataset. The site of the malignant foci with respect to sampled nodules might also give some understanding of the extent of FNA in multinodular glands. Standardisation of the pathological assessment, and sufficient follow-up across multiple centres, is required prior to the development of prediction models and modification of surgical thresholds. The descriptive parts of the main hypothesis are thus corroborated: incidental carcinoma can be found after surgery for presumed benign MNG, papillary histology is frequent (where mentioned) and the proportions vary widely between studies. The evidence is not conclusive to identify a single risk to all patients with MNG or to recommend more extensive surgery based on this risk (Chen & Haymart, 2026; Durante et al., 2023; Ito & Miyauchi, 2023).

6. Conclusion

A systematic review and meta-analysis of 11 surgical cohorts of 7,400 patients undergoing surgery for presumed benign multinodular goitre (MNG) showed a pooled proportion of incidental thyroid carcinomas after surgery of 13.7% (95% CI 7.9% to 22.7%). The most common histological type reported was papillary carcinoma, which included microcarcinoma. However, the range in both between-study heterogeneity and selection of patients for surgery precludes the use of this pooled proportion as an individual risk estimate or as proof of clinical significance of the disease when detected in any individual patient.

The results confirm that careful assessment of patients prior to surgery and counselling of the risks and benefits is important, and that there is a significant difference between nodular disease which is suitable for surgery and incidental pathology identified only following surgery. They cannot be used to routinely advocate for total thyroid surgery to avoid the "surprise" of a diagnosis of carcinoma. Varying definitions of cohorts, incomplete reporting of risk of bias elements, a lack of standardization in pathological reporting, and longer term clinical outcomes are needed to turn these postoperative detection estimates into strong recommendations for practice.

References

1. Al-Hakami, H. A., Kouther, D. A., Alsharef, J. F., Kouther, M. A., Abualola, A. H., Ghaddaf, A. A., Awad, B., & Al Garni, M. (2024). Total thyroidectomy versus partial thyroidectomy for non-toxic multinodular goiter: A systematic review and meta-analysis of randomized controlled trials. *Indian Journal of Surgical Oncology, 15*(4), 920–930. https://doi.org/10.1007/s13193-024-02057-y

2. Apostolou, K., Zivaljevic, V., Tausanovic, K., Zoric, G., Chelidonis, G., Slijepcevic, N., Jovanovic, M., & Paunovic, I. (2021). Prevalence and risk factors for thyroid cancer in patients with multinodular goitre. *BJS Open, 5*(2), zraa014. https://doi.org/10.1093/bjsopen/zraa014

3. Arican, C. D., Ozturk, T., Sager, M. S., Sertbudak, I., Teksoz, S., Saricoban, C. T., & Uygur, A. (2023). Incidental papillary microcarcinoma and papillary thyroid carcinoma in multinodular goiter. *Analytical Cellular Pathology, 2023*, Article 2768344. https://doi.org/10.1155/2023/2768344

4. Bharath, S., Yadav, S. K., Sharma, D., Jha, C. K., Mishra, A., & Mishra, S. K. (2023). Total vs less than total thyroidectomy for benign multinodular non-toxic goiter: An updated systematic review and meta-analysis. *Langenbeck's Archives of Surgery, 408*, Article 200. https://doi.org/10.1007/s00423-023-02941-1

5. Bombil, I., Bentley, A., Kruger, D., & Luvhengo, T. E. (2014). Incidental cancer in multinodular goitre post thyroidectomy. *South African Journal of Surgery, 52*(1), 5–9. https://pubmed.ncbi.nlm.nih.gov/24881131/

6. Bray, F., Laversanne, M., Sung, H., Ferlay, J., Siegel, R. L., Soerjomataram, I., & Jemal, A. (2024). Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians, 74*(3), 229–263. https://doi.org/10.3322/caac.21834

7. Campbell, M. J., Seib, C. D., Candell, L., Gosnell, J. E., Duh, Q.-Y., Clark, O. H., & Shen, W. T. (2015). The underestimated risk of cancer in patients with multinodular goiters after a benign fine needle aspiration. *World Journal of Surgery, 39*(3), 695–700. https://doi.org/10.1007/s00268-014-2854-y

8. Chen, D. W., & Haymart, M. R. (2026). Unravelling the rise in thyroid cancer incidence and addressing overdiagnosis. *Nature Reviews Endocrinology, 22*(1), 10–20. https://doi.org/10.1038/s41574-025-01168-y

9. Chen, W.-H., & Chen, C.-Y. (2022). Clinicopathologic characteristics of incidental thyroid carcinoma in euthyroid patients receiving total thyroidectomy for multinodular goiter: A retrospective cohort study. *Journal of the Chinese Medical Association, 85*(8), 839–844. https://doi.org/10.1097/JCMA.0000000000000758

10. DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188. https://doi.org/10.1016/0197-2456(86)90046-2

11. Durante, C., Hegedüs, L., Czarniecka, A., Paschke, R., Russ, G., Schmitt, F., Soares, P., Solymosi, T., & Papini, E. (2023). 2023 European *Thyroid* Association clinical practice guidelines for thyroid nodule management. European Thyroid Journal, 12(5), e230067. https://doi.org/10.1530/ETJ-23-0067

12. Fama, F., Sindoni, A., Cicciù, M., Polito, F., Piquard, A., Saint-Marc, O., Gioffrè-Florio, M., & Benvenga, S. (2018). Preoperatively undiagnosed papillary thyroid carcinoma in patients thyroidectomized for benign multinodular goiter. *Archives of Endocrinology and Metabolism, 62*(2), 139–148. https://doi.org/10.20945/2359-3997000000017

13. Gál, I., Solymosi, T., Lukács-Tóth, G., & Wéber, G. (2013). A total thyreoidectomia hatásossága és biztonságossága a jóindulatú multinodularis golyvák kezelésében [Effectiveness and safety of total thyroidectomy in the management of benign multinodular goiters]. *Magyar Sebészet, 66*(5), 245–249. https://doi.org/10.1556/MaSeb.66.2013.5.3

14. Haugen, B. R., Alexander, E. K., Bible, K. C., Doherty, G. M., Mandel, S. J., Nikiforov, Y. E., Pacini, F., Randolph, G. W., Sawka, A. M., Schlumberger, M., Schuff, K. G., Sherman, S. I., Sosa, J. A., Steward, D. L., Tuttle, R. M., & Wartofsky, L. (2016). 2015 American *Thyroid* Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: The American Thyroid Association Guidelines Task Force on Thyroid Nodules and Differentiated Thyroid Cancer. Thyroid, 26(1), 1–133. https://doi.org/10.1089/thy.2015.0020

15. Ito, Y., & Miyauchi, A. (2023). Active surveillance may be the best initial management for papillary thyroid microcarcinoma. *Journal of the Endocrine Society, 7*(7), bvad063. https://doi.org/10.1210/jendso/bvad063

16. Kaliszewski, K., Strutyńska-Karpińska, M., Zubkiewicz-Kucharska, A., Wojtczak, B., Domosławski, P., Balcerzak, W., Łukieńczuk, T., & Forkasiewicz, Z. (2016). Should the prevalence of incidental thyroid cancer determine the extent of surgery in multinodular goiter? *PLOS ONE, 11*(12), e0168654. https://doi.org/10.1371/journal.pone.0168654

17. Lin, L., & Chu, H. (2020). Meta-analysis of proportions using generalized linear mixed models. *Epidemiology, 31*(5), 713–717. https://doi.org/10.1097/EDE.0000000000001232

18. Mohamed, T. Z., Sultan, A. A. E. A., El-Din, M. T., Mostafa, A. A. E., Nafea, M. A., Kalmoush, A.-E., Nassar, M. S., Abdalgaleel, M. A., Hegab, A. M., Ibrahim, A. H., & Baheeg, M. (2022). Incidence and risk factors of thyroid malignancy in patients with toxic nodular goiter. *International Journal of Surgical Oncology, 2022*, Article 1054297. https://doi.org/10.1155/2022/1054297

19. Özpek, A., & Topal, C. S. (2019). Incidental thyroid cancer in patients undergoing surgery for benign multinodular goitre: A retrospective clinical study. *Journal of Surgical Oncology (Science Repository)*. https://doi.org/10.31487/j.JSO.2019.01.007

20. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71. https://doi.org/10.1136/bmj.n71

21. Pavone, G., Lamanna, E., Pacilli, M., Khoury, E., Ambrosi, A., & Tartaglia, N. (2026). Incidental thyroid carcinoma in surgically treated multinodular goiter: A retrospective study. *Frontiers in Surgery, 13*, Article 1825352. https://doi.org/10.3389/fsurg.2026.1825352

22. Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011).

Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ, 343*, d4002. https://doi.org/10.1136/bmj.d4002

Appendix. Database search strategies

Database-specific search expressions are given below. PubMed used exact publication-date limits, whereas Scopus and Web of Science applied publication-year restrictions.

PubMed

("multinodular goiter"[tiab] OR "multinodular goitre"[tiab] OR "nodular goiter"[tiab] OR "nodular goitre"[tiab]) AND (incidental*[tiab] OR occult[tiab] OR unsuspected[tiab] OR unexpected[tiab] OR histopatholog*[tiab] OR postoperative[tiab]) AND (carcinom*[tiab] OR malignan*[tiab] OR cancer*[tiab] OR microcarcinom*[tiab]) AND 2011/11/12:2026/11/12[dp]

Scopus

TITLE-ABS-KEY(("multinodular goiter" OR "multinodular goitre" OR "nodular goiter" OR "nodular goitre") AND (incidental* OR occult OR unsuspected OR unexpected OR histopatholog* OR postoperative) AND (carcinom* OR malignan* OR cancer*)) AND PUBYEAR > 2010 AND PUBYEAR < 2027 AND DOCTYPE(ar)

Web of Science

TS=(("multinodular goiter" OR "multinodular goitre" OR "nodular goiter" OR "nodular goitre") AND (incidental* OR occult OR unsuspected OR unexpected OR histopatholog* OR postoperative) AND (carcinom* OR malignan* OR cancer*)) AND PY=(2011-2026)