Thyroid fine-needle aspiration and the bethesda classification system

Larsen, Louise Vølund; Egset, Alice Viktoria; Holm, Camilla; Larsen, Stine Rosenkilde; Nielsen, Susanne Holm; Bach, Jacob; Helweg-Larsen, Jens Peter; Wanscher, Jens Højberg; Godballe, Christian

Thyroid fine-needle aspiration and The Bethesda Classification System

Louise Vølund Larsen[1], Alice Viktoria Egset[1], Camilla Holm[1], Stine Rosenkilde Larsen[2], Susanne Holm Nielsen[3], Jacob Bach[4], Jens Peter Helweg-Larsen[5], Jens Højbjerg Wanscher[1] & Christian Godballe[1]

ABSTRACT

INTRODUCTION: Fine-needle aspiration (FNA) is a cornerstone in diagnosing thyroid nodules. For decades, Danish FNA have been categorised into the groups: "FNA not performed", "Inadequate", "Cystic", "Inconclusive", "Benign", "Suspicious", "Malignant" and "Information missing". Internationally, The Bethesda Classification System (TBCS) is increasingly accepted, especially owing to a detailed specification of FNA suspicious for malignancy. The Danish "Suspicious" group is very broad and includes atypia, follicular neoplasia and FNA suspicious of other malignancies. The purpose of this study was to apply TBCS to the Danish "Suspicious" FNA group and to estimate the frequency of malignancy in the individual Bethesda groups (BG).
METHODS: This descriptive study is based on a prospective cohort from the THYKIR database. It includes 479 patients with a "Suspicious" FNA and surgical treatment in The Region of Southern Denmark from 2001 to 2013. Based on pathology records, FNA was classified according to the TBCS. Malignancy was determined by the histological diagnosis from the THYKIR database.
RESULTS: The Danish "Suspicious" group was allocated to the BG I, II, III, IV, V and VI with a malignancy risk of 36.4%, 13.3%, 17.2%, 16.1%, 55.3% and 88.2%, respectively.
CONCLUSIONS: The Danish "Suspicious" group contains a broad spectrum of BG with varying malignancy risk. The results indicate a need for standardisation of the Danish FNA classification. A national introduction of the TBCS might secure an international and comparable standard.
FUNDING: none.
TRIAL REGISTRATION: not relevant.

Thyroid carcinoma is the most frequent endocrine malignancy in Europe [1]. In 2015, The Danish Thyroid Cancer Database (DATHYRCA) registered 312 patients with thyroid cancer [2] in the Danish population of 5.6 million people [3]. A Danish study showed that 67% of Danish thyroid carcinomas are of the papillary type, 18% of the follicular type, 7% of the medullary type and 8% of the undifferentiated (anaplastic) type [4].

Palpable thyroid nodules are frequent with a prevalence of 5% in women and 1% in men, but only 7-15% of thyroid nodules are malignant [5]. Therefore, it is important to have an accurate preoperative diagnostic tool capable of discriminating between benign and malign

thyroid nodules to reduce any unnecessary surgery. Fine-needle aspiration (FNA) is a cornerstone in the diagnostic strategy for thyroid nodules and is recommended by The American Thyroid Association (ATA) guidelines because it is the most precise and cost-effective method for evaluation of thyroid nodules [5].

In a Danish study, the sensitivity and specificity for malignancy were 73.9% and 99.2%, respectively [6]. The positive predictive value was 89.5% and the negative predictive value was 97.7% [6].

Previously, there was no agreement on the number, types or terminology of FNA categories, which made it difficult to compare results between centres and nations. In addition, differences between pathologists and clinicians in their understanding of pathology reports have been shown [7]. This may partly be due to vague descriptions such as "indeterminate" or "cannot exclude", which can lead to confusion. To resolve these problems, an increasing number of centres have implemented The Bethesda Classification System (TBCS), which is the offspring of the 2007 National Cancer Institute Thyroid Fine Needle Aspiration State of the Science Conference. The aim of TBCS is to provide consensus recommendations for diagnostic terminology and morphological criteria to increase the quality and reproducibility of thyroid cytology reporting [8]. TBCS is recommended by the 2015 ATA guidelines [5].

TBCS consists of six diagnostic categories: I Nondiagnostic/Unsatisfactory; II Benign; III Atypia of undetermined significance/Follicular lesion of undetermined significance (AUS/FLUS); IV Follicular neoplasm/Suspicious for follicular neoplasm (FN/SFN); V Suspicious for malignancy (SFM) and VI Malignant [8].

In Denmark, thyroid cytology is categorised into eight groups as defined in The Danish Thyroid Surgery Database (THYKIR): "FNA not performed", "Inadequate", "Cystic", "Inconclusive", "Benign", "Suspicious", "Malignant" or "Information missing". The Danish "Suspicious" group is very broad and includes follicular neoplasia, atypia and FNA suspicious for other malignancy. A recent Danish study found the frequency of malignancy in the Danish "Suspicious" FNA group to be 22.4% [9].

A competent classification system for FNA is im-

FIGURE 1

Flow chart of the study population.
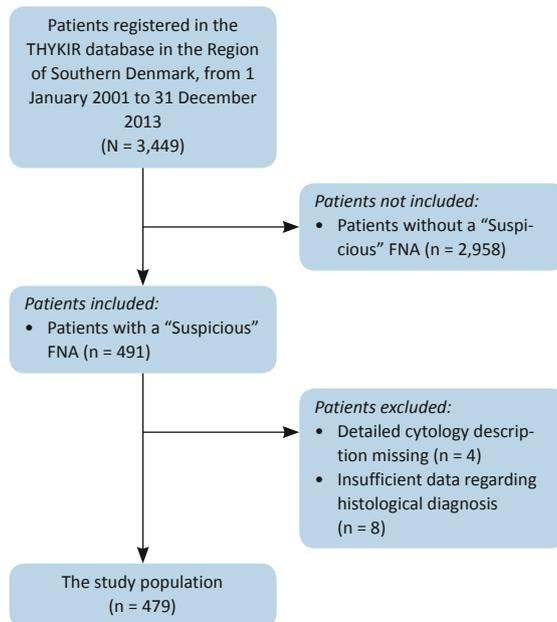
FNA = fine-needle aspiration; THYKIR = The Danish Thyroid Surgery Database.

portant because of its role in treatment decision. Due to the broad nature of the Danish "Suspicious" group, it is of great interest to investigate the distribution of FNA results and the malignancy risk in each Bethesda group (BG) when TBCS is applied to the Danish "Suspicious" group. A classification of the Danish "Suspicious" group based on TBCS has never been done. The purpose of this study was to apply TBCS to the Danish "Suspicious" FNA group and to estimate the frequency of malignancy in each individual BG.

METHODS

This is a descriptive study based on a prospectively registered cohort from the THYKIR database, which contains data on thyroid surgery performed in Denmark. Among the 3,449 patients who had thyroid surgery performed at the departments of Otorhinolaryngology, Head & Neck Surgery in the Region of Southern Denmark (Odense University Hospital, Vejle Hospital, Esbjerg Hospital and Soenderborg Hospital) from 1 January 2001 to 31 December 2013, 491 patients had a "Suspicious" FNA and were included in this study. The patients were linked by their personal identity number to the National Pathology Register (PATOBANKEN) to identify their pathological records.

Of the 491 patients, 12 patients were excluded; four due to lack of cytological descriptions of FNA and eight due to insufficient data concerning their histo-

logical diagnosis. The final study population consisted of 479 patients (**Figure 1**).

Cytological descriptions of FNA from PATOBANKEN were retrospectively analysed and classified according to TBCS without knowledge of the histological diagnosis. In case of uncertainty about categorisation, a team consisting of two clinicians and one pathologist made a final consensus-based decision. This was the case in 48 of 479 (10%) cases. If the team could not reach an agreement, the pathologist made the final decision.

When the patient had more than one FNA, the one with the highest ("most malignant") BG was chosen. Pathologists at university hospitals see more FNAs than pathologists at smaller hospitals and it was considered reasonable to assume that pathologists at university hospitals had more expertise. Therefore, if the FNA had been revised by a university hospital, their diagnosis was chosen regardless if it was a higher or lower BG than the original diagnosis.

The frequency of malignancy was subsequently determined in each BG by the histological diagnosis extracted from the THYKIR database. Data were stored, processed and analysed anonymously in an Excel file within SharePoint. The study was approved by the Danish Data Protection Agency/Region of Southern Denmark (record no.:15/36370 and 15/28871). The information was identified from already approved databases, and no patient records were used.

Trial registration: not relevant.

RESULTS

The study population included 479 patients; 371 (77.5%) women and 108 (22.5%) men. The age ranged from 16 to 91 years with a median age of 53 years at the time of surgery. The FNAs were classified into six groups according to TBCS, and the frequency of malignancy in each BG was determined based on the histological diagnosis. The results are presented in **Table 1**. A total of 43 (9.0%) of the patients had an FNA classified into BG I, II or VI. The most dominant group was BG IV followed by BG III, and combined these groups contained 398 (83.1%) patients.

DISCUSSION

This is the first study addressing the distribution of Danish FNA according to TBCS and determining the frequency of malignancy in each BG for the Danish "Suspicious" group. TBCS, first published by Cibas et al in 2009, contained estimates of the cancer risk in each BG based on a literature review and expert opinion [8]. Many centres have published their frequency of malignancy after adopting TBCS and, despite variability, their results for BG IV-VI were overall comparable with the range estimated by Cibas et al [8, 10, 11].

The precise risk of malignancy in group I-III is difficult to assess from the literature. The studies often only included patients with surgical follow-up, and excluded those without. The patients who underwent surgery would presumably have other signs of malignancy leading to surgery making the risk of overestimating the malignancy risk in BG I-III substantial.

Our study found a higher frequency of malignancy in BG I (36.4%) and II (13.3%), and a lower frequency in BG VI (88.2%) compared with Cibas et al, where the risk of malignancy was estimated to 1-4%, 0-3% and 97-99%, respectively [8]. This was most likely due to the fact that our study did not include patients with a FNA classified as "Inadequate", "Cystic", "Inconclusive", "Benign" or "Malignant". Our study may therefore have overestimated the frequency of malignancy in BG I and II, and underestimated malignancy in BG VI. Hence, only our results from BG III, IV and V can truly be compared with the literature.

In our study, the frequency of malignancy for BG III was 17.2%, which was higher than the 5-15% estimated by Cibas et al [8]. Among other studies, the results for the BG III have shown variability, and many have implied that the malignancy risk is greater than estimated buy Cibas et al [8, 10-12]. Ho et al found the frequency of malignancy to be 26.6-37.8% [12]; a review by Ohori et al found a mean malignancy of 26.3% [11], while a meta-analysis of Bongiovanni et al only showed a slightly higher risk of 15.9% [10] than Cibas et al [8]. As mentioned, not all patients in BG III had surgical follow-up, and the variation in the literature might be due to the use of different criteria for selecting patients for operation. If it is assumed that all Danish patients with AUS/FLUS (BG III) are categorised as "Suspicious", almost all are surgically treated, and the risk of an overestimation is limited. However, as our study did not evaluate all Danish FNA, it is possible that Danish groups such as the "Inconclusive" group may contain some AUS/FLUS, making our BG III a selective group with a risk of overestimating the malignancy risk. This might also explain why our frequency of malignancy, surprisingly, was higher in BG III than in BG IV as it is assumed that all FN/SFN (BG IV) are categorised as "Suspicious" and operated. Therefore, the malignancy risk is most likely not overestimated in our BG IV.

Another problem is the tendency to overuse BG III. It is recommended that BG III is used only as a last resort and to limit the use to ≤ 7% of all FNA [8]. In our study, 18.2% were classified as BG III, but this cannot be compared with the limit of 7% as this is only the distribution in the "Suspicious" group.

Most of the patients (64.9%) with a "Suspicious" FNA were classified into BG IV, and 16.1% of these had a malignant histology. This is in the lower end of the 15-

30% that Cibas et al suggested [8] and much lower than reported by Bongiovanni et al (26.1%) [10] and Ohori et al (26.4%) [11].

The frequency of malignancy in BG V (55.3%) was lower than both the 60-75% estimated by Cibas et al [8], and the frequencies found by Bongiovanni et al (75.2%) [10] and Ohori et al (79.1%) [11].

The differences between our results for BG III-V and the literature might be explained by a different distribution in Denmark, but might also be due to misclassification because the cytological descriptions were not made with subsequent reclassification into TBCS in mind. Furthermore, an FNA classified as "Suspicious" resulted in the same treatment regardless if it was described as AUS, FN, SFN or SFM. Pathologists might have been more precise in distinguishing borderline cases if the treatment of these had been different. Due to the relatively low dietary iodine intake in Denmark, the national diagnostic strategy for thyroid nodules include thyroid scintigraphy. This is not the case in the US or in most European countries. Therefore, direct comparisons of BG results have to be interpreted with caution.

Combined, the malignancy risk in the Danish "Suspicious" FNA group was 22.4% [9]. Our study showed that if "Suspicious" FNA are classified into TBCS, the malignancy risk was much higher in BG V than in BG III and IV. This indicates that patients would get a more accurate risk estimation by using TBCS which is important to both the patients and the doctors involved. However, we do not expect significant changes in the choice of surgical procedure, i.e. hemi-throidectomy versus total thyroidectomy. However, a more precise knowledge about the risk of malignancy may influence the surgical setup concerning the availability of frozen section histology and the experience of the surgical team. Furthermore, a more precise communication between the pathologist and clinicians will benefit the patient.

TABLE 1

Distribution of fine-needle aspirations in the Danish "Suspicious" group when classified according to The Bethesda Classification System, and the frequency of malignancy in each Bethesda Group.

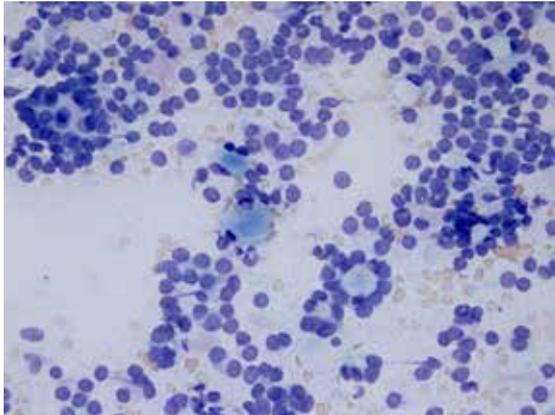| Bethesda Group	n (%[a])	Malignant cases, n (%[b])
I: Non-diagnostic/unsatisfactory	11 (2.3)	4 (36.4)
II: Benign	15 (3.1)	2 (13.3)
III: Atypia of undetermined significance/follicular lesion of undetermined significance	87 (18.2)	15 (17.2)
IV: Follicular neoplasm/suspicious for follicular neoplasm	311 (64.9)	50 (16.1)
V: Suspicious for malignancy	38 (7.9)	21 (55.3)
VI: Malignant	17 (3.5)	15 (88.2)
Total	479 (100)	107 (22.3)

a) % of the 479 patients
b) % of malignant cases in each Bethesda Group

Fine-needle aspiration with follicular neoplasm.

Cytological descriptions may be unclear, making the clinician insecure and perhaps more prone to resort to surgery than surveillance. By forcing the pathologist to give a precise Bethesda classification this may be prevented.

Patients with FNA classified as "Suspicious" were anticipated to be distributed among BG III, IV and V, but 9% were classified in BG I, II or VI. This implies that the Danish "Suspicious" group cannot be compared uncritically with BG III-V from other centres. This may be due to different diagnostic and morphological criteria in the two classification systems and indicates the importance of having the same classification systems across centres and nations to be able to compare data. However, the most important reason for the 9% classified as BG I, II and VI is presumably miscommunication between the pathologist and the clinicians, as it were the clinicians who initially interpreted the cytological description and allocated the FNA to the "Suspicious" group in the THYKIR database. In some cases, the cytological description was diffuse and without a clear conclusion, making an unambiguous registration very difficult. Therefore, the initial interpretation performed by the registering clinician and the interpretation performed by our team may differ significantly. This very strongly implies the need for a standardised reporting system, such as TBCS, to improve communication among pathologists and clinicians.

An important limitation to this study was that the FNA were classified into TBCS categories based on interpretation of cytological descriptions and not directly by the pathologist performing the cytological examination. Another limitation was that data were accepted as reported to the THYKIR database. This is an issue because errors in the reporting process may occur, which might also be part of the explanation for some of the 9% classified as BG I, II or VI. Furthermore, accepting the histological diagnosis as the correct diagnosis to the FNA could be problematic because the study design did not guarantee that the nodule biopsied was the same as the one containing a malignant diagnosis.

Among the strengths in this study were the prospective data collection from both THYKIR and PATO-BANKEN, the coverage of a specific geographical referral area (The Region of Southern Denmark) and that classification of the FNA into TBCS occurred without knowledge of the histological diagnosis. Furthermore, the usage of the THYKIR database with a completeness of around 97% [13] and the ability to find all the patients' pathology records via their personal identity number further strengthen the study. The mentioned strengths reduce the risk of bias and increase the generalisability of the study.

This study indicates the frequency of malignancy in each BG for the Danish "Suspicious" group. To determine the exact distribution of malignancy, further studies are needed to investigate, e.g., the frequency of malignancy after implementing TBCS or to reexamine the FNA instead of using pathological records.

CONCLUSIONS

The Danish "Suspicious" FNA group contains a broad spectrum of BG with varying malignancy risk. The malignancy risk for the BG I, II, III, IV, V and VI was 36.4%, 13.3%, 17.2%, 16.1%, 55.3% and 88.2%, respectively. The results indicate a need for standardisation of the Danish FNA classification. A national introduction of TBCS might secure an international and comparable standard.

LITERATURE
1. van der Zwan JM, Mallone S, van Dijk B et al. Carcinoma of endocrine organs: results of the RARECARE project. Eur J Cancer 2012;48:1923-31.
2. Eriksen JG, Jovanovic A, Johansen J et al. Årsrapport 2015 for den kliniske kvalitetsdatabase DAHANCA. Copenhagen: DAHANCA, 2016.
3. Statistics Denmark. Denmark in figures 2015. Copenhagen: Statistics Denmark, 2015. http://hdl.handle.net/109.1.3/9020c5fa-90c1-4cb9-bdad-796e69c9c878 (23 May 2017).
4. Londero SC, Krogdahl A, Bastholt L et al. Papillary thyroid carcinoma in Denmark 1996-2008: an investigation of changes in incidence. Cancer Epidemiol 2013;37:e1-e6.
5. Haugen BR, Alexander EK, Bible KC et al. 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 2016;26:1-133.
6. Rossing M, Nygaard B, Nielsen FC et al. High prevalence of papillary thyroid microcarcinoma in Danish patients: a prospective study of 854 consecutive patients with a cold thyroid nodule undergoing fine-needle aspiration. Eur Thyroid J 2012;1:110-7.
7. Powsner SM, Costa J, Homer RJ. Clinicians are from Mars and pathologists are from Venus. Arch Pathol Lab Med 2000;124:1040-6.
8. Cibas ES, Ali SZ. The Bethesda System for Reporting Thyroid Cytopathology. Thyroid 2009;19:1159-65.
9. Egset AV, Holm C, Larsen SR et al. Risk of malignancy in fine-needle aspiration biopsy in patients with thyroid nodules. Dan Med J 2017; 64(2):A5320.
10. Bongiovanni M, Spitale A, Faquin WC et al. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. Acta Cytol 2012;56:333-9.

11. Ohori NP, Schoedel KE. Variability in the atypia of undetermined significance/follicular lesion of undetermined significance diagnosis in the Bethesda System for Reporting Thyroid Cytopathology: sources and recommendations. Acta Cytol 2011;55:492-8.
12. Ho AS, Sarti EE, Jain KS et al. Malignancy rate in thyroid nodules classified as Bethesda category III (AUS/FLUS). Thyroid 2014;24:832-9.
13. Godballe C, Madsen AR, Pedersen HB et al. Post-thyroidectomy hemorrhage: a national study of patients treated at the Danish departments of ENT Head and Neck Surgery. Eur Arch Otorhinolaryngol 2009;266:1945-52.