



Next Generation Quality: Assessing Diagnostic Rate Variance with Funnel Plots in 10,540 Thyroid Fine Needle Aspiration Biopsies

Chandni Jain¹, Gaurav Jain², Michael Bonert³, Asghar Naqvi³, Mohamed Khodary³, Ipshita Kak³, Michael Gupta⁴
¹McMaster University ²University of Ottawa ³Division of Pathology and ⁴Otolaryngology Head and Neck Surgery at St. Joseph’s Healthcare Hamilton



Background and Objectives

Continuous quality management using objective clinical data to provide ongoing feedback to healthcare providers has traditionally been a time consuming and data intensive process. Automated extraction and analysis using funnel plots may offer a more efficient way to compare diagnostic rates (DRs) and infer statistically significant DR variation in the healthcare setting for quality improvement purposes.

The objective of this study was to use a novel technique of automated data extraction using a custom-built automated system and funnel plot analysis to compare the diagnostic rates for thyroid fine needle aspiration biopsies among head and neck pathologists at a tertiary care centre to detect bias and for quality improvement purposes. The correlation between TI-RADS score and malignancy risk at our centre was also assessed.

Design / Methods

A custom computer program retrieved and categorized information from all thyroid fine needle aspiration (FNA) biopsy reports for cases reported between 2011- 2019 at an academic tertiary care centre. DRs for malignant, suspicious for malignancy, follicular neoplasm, follicular lesion/atypia of undetermined significance (FLUS/AUS), benign and unsatisfactory diagnoses were calculated for all pathologists reporting >100 cases and graphed on funnel plots centered on the group median diagnostic rate (GMDR) with outer bounds defined by the boundaries of p=0.05 and p=0.001. Thyroid image reporting and data system (TI-RADS) score was extracted where this was provided and correlated with the cytologic diagnosis.

Inclusion Criteria

Thyroid cytopathology reports from St. Joseph’s Healthcare Hamilton and Hamilton Health Sciences Hospitals from 2011-2019

Exclusion Criteria

Pathologists with <100 thyroid FNA cytopathology reports

Results

10,540 thyroid FNAs were retrieved and >97% of cases could be coded by the program. Accuracy of the automated categorization was 98% based on 200 pathologist audited cases. 28 pathologists reported between 104 and 827 cases each. The GMDR (range) for the six diagnoses were: 5% (1-15%) for malignant, 2% (0-7%) for suspicious, 6% (1-16%) for follicular neoplasm, 14% (1-20%) for FLUS/AUS, 42% (29-55%) for benign and 35% (19-43%) for unsatisfactory.

Disclosure: No financial conflicts to declare

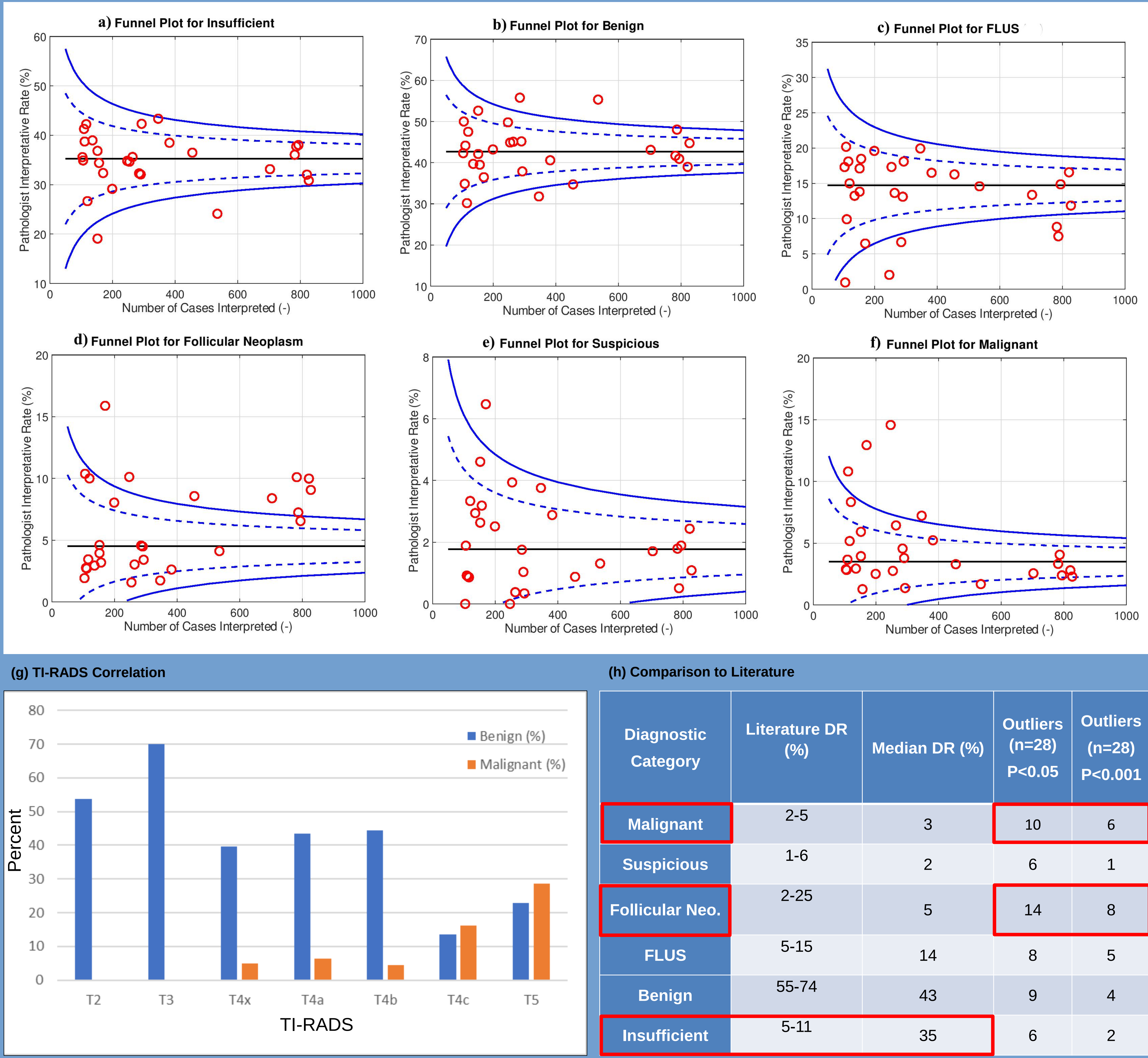


Figure 1 (a-f): The red circles represent individual pathologist. The black lines are the group median diagnostic rates. The dashed blue lines represent the 95% confidence interval (CI) and the solid blue lines are the 99.9% CI; Diagnostic rates (DR) between the two lines are likely variations due to sampling bias, versus DRs outside the solid blue line are likely due to interpretive bias (p<0.001). Figure 1 (g): shows the correlation between TI-RADS score and FNA malignant diagnosis. (h) MGCL by diagnostic category.

Results

Pathologists outside the 99.9% confidence interval (p<0.001) were: 6 for malignant, 1 for suspicious, 8 for follicular neoplasm, 5 for FLUS/AUS, 4 for benign and 2 for unsatisfactory diagnoses. TI-RADS score was available for 699 cases and the percentage reported malignant for TI-RADS 2, 3, 4A, 4B, 4C, and 5 were 0%, 0%, 6.5%, 4.4%, 16.2%, and 28.6% respectively. Overall, there is uniformity in the majority of DR for thyroid FNAs, with evidence of diagnostic bias in each Bethesda category among pathologists at our institution, confirmed over a large sample size. TI-RADS score correlates well with thyroid cytopathology diagnosis.

Limitations

This study was bound by a limited ethics approval. As such, individual pathologists could not be identified to investigate reasons for any discrepancies. Furthermore, automated extraction could not differentiate between Kwak and ACR TI-RADS scores from the pathology requisition, and the current system could not link ultrasound reports for a larger sample size.

Implications and Future Directions

This is the first study to use Funnel Plots for quality management in Head and Neck surgical pathology. Automated extraction and categorization from free text pathology reports can be used to monitor pathologist diagnostic rates using large samples to minimize sampling bias. This is an efficient and feasible method to identify areas in need of improvement as part of a comprehensive data driven quality management program. We plan to follow diagnostic rates and TI-RADS pathology correlation at our centre. There is need for further investigation and targeted improvements regarding the identified lack of consistency of follicular neoplasm and malignancy diagnostic rates at our centre and higher insufficient diagnosis compared to the literature.

Conclusions

Automatic extraction and categorization of diagnostic information from free text pathology reports efficiently yields aggregate data that can be used for continuous quality management. Utilization of funnel plots for quality improvement purposes is a novel way to interpret large amounts of data and identify statistically and clinically significant outliers. At our centre there is room for improvement in FNA diagnostic categories, with the greatest variability in DR seen in malignant and follicular neoplasm. This can have significant surgical implications. The correlation between TI-RADS score and malignancy diagnosed on FNA in these preliminary results was reassuring and comparable with the literature. There is a need for further investigation and targeted improvement of follicular neoplasm, malignant, and nondiagnostic results at our centre.