

OPEN

Application of Next Generation Quality/Statistical Process Control and Expert-Led Case Review to Increase the Consistency of Diagnostic Rates in Precancerous Colorectal Polyps

Michael Bonert, MD, MASc, FRCPC; Andrew Collins, MD; Ted Xenodemetropoulos, MD, MSc, FRCPC; Jennifer M. Dmetrichuk, PhD, MBChB, FRCPC; Sahar Al-Haddad, MBChB, FRCPC; Pierre Major, MD, FRCPC; Asghar Naqvi, MD, FRCPC

Background: Prior work suggests high interrater variability in the pathologist diagnostic rate (PDR) of the precancerous polyp sessile serrated adenoma (SSA). **Objectives:** To improve the diagnostic consistency in the pathological evaluation of colorectal polyp specimens with diagnostic rate awareness, using funnel plots (FPs)/control charts (CCs), and a focused group case review. **Methods:** All colorectal polyp specimen (CRPS) reports September 2015 to August 2017 were analyzed at one institution. PDRs were extracted using a hierarchical free-text string matching algorithm and visualized using FPs, showing pathologist specimen volume versus PDR, and CCs, showing pathologist versus normed PDR. The FPs/CCs were centered on the group median diagnostic rate (GMDR). Pathologists were shown their baseline SSA diagnostic rate in relation to the practice, and in January 2017, there was a focused group case review/open discussion of approximately 40 sequential cases signed as SSA with a gastrointestinal pathology expert. **Results:** Nine pathologists interpreted more than 250 CRPSs per year. FPs/CCs for the first and second years showed 6/4 and 3/1 $P < .05/P < .001$ pathologist outliers, respectively, in relation to the GMDR for SSA and 0/0 and 0/0 $P < .05/P < .001$ pathologist outliers, respectively, in relation to the GMDR for tubular adenoma (TA). An in silico kappa (ISK) for SSA improved from 0.52 to 0.62. **Conclusion:** Diagnostic rate awareness facilitated by FPs/CCs coupled with focused expert-led reviews may help calibrate PDR. Variation in SSA PDRs still remains high in relation to TA. ISK represents an intuitive, useful metric and Next Generation Quality/Statistical Process Control a promising approach for objectively increasing diagnostic consistency.

Key words: colorectal polyps, continuous quality improvement, pathologist diagnostic rate, sessile serrated adenoma, Statistical Process Control

Colonoscopy is routinely used in the screening, surveillance, and diagnosis of colorectal cancer and potentially premalignant colorectal polyps. Among colorectal biopsies, there is a pathological spectrum

ranging from normal mucosa to hyperplastic polyps, sessile serrated adenomas (SSAs), dysplastic polyps (eg, SSAs with dysplasia), and carcinoma. SSAs represent a transition in this serrated carcinogenesis pathway from normal mucosa/hyperplastic polyps, which are considered to be benign lesions with

Author Affiliations: Departments of Pathology and Molecular Medicine (Drs Bonert, Dmetrichuk, Al-Haddad, and Naqvi), Medicine (Dr Collins), and Oncology (Dr Major), McMaster University, Hamilton, Ontario, Canada; St. Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada (Drs Bonert and Naqvi); Farncombe Family Digestive Health Research Institute/Division of Gastroenterology, Department of Medicine, McMaster University, Hamilton, Ontario, Canada (Dr Xenodemetropoulos); and Juravinski Hospital/Hamilton Health Sciences Centre/Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, Canada (Ms Al-Haddad and Dr Major).

Correspondence: Michael Bonert, MD, MASc, FRCPC, Department of Pathology, St. Joseph's Healthcare Hamilton, Room L206, 50 Charlton Ave East, Hamilton, ON L8N 4A6, Canada (bonertm@mcmaster.ca).

Earlier versions of this work were presented at the United States and Canadian Academy of Pathology Annual Meeting 2018 in Vancouver, Canada, and the Canadian Digestive Disease Week 2019 in Banff, Canada. The Research Institute of St. Joe's Hamilton provided funding support for publication. The authors acknowledge participating pathology staff at St Joseph's Healthcare Hamilton.

Authors' contributions: M.B. and A.N. conceived the study. M.B. did the programming and data analysis. M.B. and A.C. wrote the first draft of the manuscript. J.M.D. audited computer categorizations from an early version of the program. S.A. is a gastrointestinal pathologist and led the case review. A.N. and M.B. audited the computer categorizations. All

authors interpreted the data and critically reviewed the manuscript. All authors approved the final manuscript.

The data sets generated and/or analyzed during the current study are not publicly available due to confidentiality reasons but aggregate data are available from the corresponding author on request.

The first author (M.B.) retains the copyright on the computer code that was written outside of his employment relationship with McMaster University/St. Joseph's Healthcare Hamilton/Hamilton Regional Laboratory Medicine Program. The other authors report no conflict of interest.

This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's Web site (www.qmhcjournal.com).

Q Manage Health Care
Vol. 30, No. 3, pp. 176–183

Copyright © 2021 The Authors. Published by Wolters Kluwer Health, Inc.
DOI: 10.1097/QMH.0000000000000299

no malignant potential, to SSAs, which are known precursors of colorectal adenocarcinoma.¹ The true prevalence of SSA remains unclear; literature values suggest they represent between 2.8% and 9% of colorectal polyps.^{2–5} In the context of serrated polyps, SSAs are thought to account for approximately 20% of serrated colorectal polyps.^{6–8}

Given the risk of progression to adenocarcinoma, accurate endoscopic identification and resection, as well as an accurate histological diagnosis of SSAs, are critical for optimal patient care.¹ Inconsistencies have an impact on patient risk assessment and recommended surveillance intervals.^{9,10} For instance, recommended follow-up for hyperplastic polyps is 10 years, whereas for SSAs varies between 3 and 5–10 years depending on size, number, and histology.^{9,10} Pathological overcall of an adenoma therefore may lead to shortened surveillance intervals, subsequent procedural risks, inconvenience, and stress to the patient, poor resource utilization, and thus increase in health care costs.^{11,12} On the other hand, undercalling premalignant colorectal polyps may result in inappropriate surveillance intervals, with a risk of undetected dysplastic progression and a potential impact on overall morbidity and mortality.

Despite significant clinical and economic implications of an inappropriate SSA diagnosis, pathologist agreement is moderate compared with the diagnosis of tubular adenoma (TA). Interobserver agreement rates for the histopathological diagnosis of SSAs have κ values ranging from 0.16 to 0.38.^{13–15} In 2007, Glatz et al¹⁶ examined the diagnostic variability among 168 international pathologists using representative images of hematoxylin and eosin–stained sections showing 20 colorectal polyps. Interobserver variability was most pronounced for SSAs (54% correct compared with 90% for TA), which were most often misdiagnosed as hyperplastic polyps or traditional serrated adenomas. Pathologists with gastrointestinal subspecialty training and those who had read a reference article gave a significantly higher percentage of correct answers for SSAs.¹⁶ Khalid et al¹³ showed that in a group of 40 previously diagnosed hyperplastic polyps, 85% were reinterpreted as SSAs by gastrointestinal pathologists; however, this was with a poor κ value (0.16) among the specialized gastrointestinal pathologists. Because of the diagnostic challenges, a significant effort has been made to establish universal criteria for SSAs. However, despite consensus papers on diagnostic criteria of colorectal lesions,^{17,18} the criteria may not be consistently applied.

We previously quantified the SSA rate in our institution and noted that the diagnostic rates varied significantly (unpublished observations), likely due to diagnostic bias. If true, the diagnostic rate could potentially be corrected through education, and we sought to achieve this as part of a quality improvement initiative.

Funnel plots (FPs) can be used to assess bias; they have been used in meta-analyses to detect publication bias and in quality of care.¹⁹ They have been used to assess cesarean section rates between institutions²⁰

and surgeons.²¹ We applied them in anatomical pathology to assess variation in prostate biopsies²² and immunostain use in relation to the diagnostic category in prostate biopsies.²³ In large-volume practices, FPs can detect relatively small significant differences that would otherwise be resource-intensive to identify via traditional case review.

FPs are similar to control charts (CCs), which are widely used in manufacturing engineering in Statistical Process Control (SPC). SPC first appeared in medical literature in 1966²⁴ and is also termed Next Generation Quality (NGQ).^{25,26} NGQ is continuous quality improvement using objective clinical data to provide feedback within a formal setting. In pathology, the feedback loop, from a process perspective, has been relatively weak; pathologists generally do not know their diagnostic call rates. The use of SPC in health care is outlined in a large systematic review published in 2017, which showed that SPC can indeed be a powerful and versatile tool for quality improvement in health care.²⁷

Overall, agreement among pathologists with regard to the pathological diagnosis of SSAs is poor, leading to variability in call rates and a multitude of clinical, economic, and psychological implications. To our knowledge, there has been no attempt to implement SPC within anatomical pathology as a means to improve agreement in the diagnosis of SSA. The aim of this study was to evaluate interrater variability in the pathologist diagnostic rate (PDR) of SSA (in relation to TA) at our institution using FPs and to assess the impact of PDR awareness and focused expert-led review on interrater variability as a potential method of quality improvement.

METHODS

Ethics

Research ethics board approval (HiREB 2016-2295-C; HiREB 2018-4445-C) with sign off from the laboratory director was obtained to access pathology and colorectal polyp reports. Consent from the involved health care providers was obtained as per the ethical framework provided by the World Health Organization in “Ethical issues in Patient Safety Research.”²⁸

Data collections and analysis

The data were extracted from the Laboratory Information System via a structured text dump. The pathology reports were transferred to a fully encrypted computer and anonymized of patient identifiers with a custom program. The files with patient identifiers were deleted with a secure delete.

Subsequently, pathology reports with colorectal polyp specimens were retrieved from the de-identified files with the following search terms:

1. “polyp” within the “source of specimen” section of the report;
2. One of the following words: “colon,” “rectum,” “rectal,” “cecum,” “cecal,” “rectosigmoid” in the “source of specimen” section of the report.

The pathology reports were then separated into parts (using the “source of specimen” section) such that individual colorectal polyp specimens could subsequently be retrieved and analyzed.

The colorectal polyp specimens were classified with a hierarchical free-text string matching algorithm (HFTSMA) that uses ^a(1) dictionary of diagnostic terms/categories and (2) a hierarchy of categories. The dictionary and hierarchy are found in Supplemental Digital Content (SDC) Appendix A and Appendix B (available at: <http://links.lww.com/QMH/A51>).

The HFTSMA was previously described in a prior article.²² Briefly:

1. The diagnosis section of the colorectal polyp specimen was searched for diagnostic terms and if 1 or more terms were found (based on an exact string match or fuzzy string matching using the library “google-diff-patch-match”), the relevant diagnostic code was assigned.
2. The hierarchy was applied; this selectively removes diagnostic codes that are deemed mutually exclusive to others within a colorectal polyp specimen (e.g. “suspicious of adenocarcinoma” supercedes “adenocarcinoma”).
3. The miss rate (noncoded cases) was counted.
4. A random sample of the (fully anonymized) study set was selected and audited to assess the coding accuracy; this was used to revise (1) the dictionary of diagnostic terms/categories and (2) the hierarchy of categories.
5. Steps 1 to 4 are repeated until the miss rate and coding accuracy were deemed acceptable.

Before analysis work was done, the remaining (non-patient) identifiers were replaced by anonymous identifiers and the anonymous identifiers linked to the plain text in a file separate from the data. The fully anonymized results were then tabulated.

With the consent of the pathologists, their baseline diagnostic rate data were decoded and displayed to them individually, using FPs (centered on the group median diagnostic rate) that showed the individual’s diagnostic rate in relation to the anonymized rates of the other pathologists reading more than 250 colorectal polyp specimens per year.

The FPs were created with custom code that was previously described.^{22,23} The group median rate was chosen as the center line of the funnel. The funnel lines were calculated with the normal approximation to the binomial distribution. The *P* values follow directly from the confidence interval; there is a 5% chance (*P* = .05) of falling outside the 95% confidence interval and a 0.1% chance (*P* = .001) of falling outside of the 99.9% confidence interval.

As per the study protocol, the individual pathologist could request (in writing) the data (surgical numbers, diagnostic text, computerized coding) for the specimens they interpreted for review.

An audit of 400 randomly selected colorectal polyp specimens was assessed by pathologist (A.N.) to determine whether the hierarchical HFTSMA improperly categorized specimens.

Expert-led case review

During year 2, to complement the data analysis, an expert-led review was done with approximately 40 sequential SSA cases. The focus of the session was the histomorphology of SSAs. The expert was blinded to the size, clinical history, and other pathology specimens from the patient. This followed brief one-on-one meetings in which the polyp diagnostic rates were disclosed.

The staff pathologists at our institution met for this purpose as a group with a gastrointestinal pathologist (S.A.). The review consisted of the gastrointestinal pathologist assessing the case, followed by verbally garnering each pathologist’s diagnostic opinion. After all of the nongastrointestinal pathologists had offered their opinion, the gastrointestinal pathologist gave her diagnostic opinion with its rationale. The narrow focus of the exercise was to determine “Is this a sessile serrated adenoma?” If it is, “Why?” and if it is not, “Why not?”

In silico kappa

In silico kappa (ISK) (λ) values were obtained by computer simulation. This involved the following: ⁽¹⁾ generating a random set of specimens; ⁽²⁾ “interpreting” the random set of specimens, using the diagnostic rates of the pathologists in the study; ⁽³⁾ calculating Fleiss’ kappa (based on the interpretations); and ⁽⁴⁾ repeating steps (1)-(3) with larger and larger sets of specimens until the kappa value was deemed converged. The calculation makes use of a “maximal diagnostic overlap assumption”; this assumption is necessary as the diagnostic rates are insufficient to infer the amount of diagnostic overlap. Conceptually, ISK is related to “kappa max” described by Sim and Wright.²⁹

Separate ISKs were calculated using (1) the (raw) diagnostic rates (as in the FPs), and (2) the normalized diagnostic rate (as found in the CCs). Details of the calculation are found in SDC Appendix D (available at: <http://links.lww.com/QMH/A51>).

We have used lambda (λ)—the letter that follows kappa (κ) in the Greek alphabet—to denote the simulated kappa—such that it should not be confused with a (traditional) kappa generated by pathologist interpretations. An ISK generated with the normalized diagnostic rate (as found in a CC) was call normed in silico kappa, abbreviated as NISK.

RESULTS

Colorectal polyp specimens

A total of 7054 colorectal polyp specimens in the study period were retrieved. The first and second years had 3656 and 3398 colorectal polyp specimens, respectively. All pathologists reading more than 250 colorectal polyp specimens per year were interested in knowing their rates and consented to seeing their rate in relation to the anonymized rates of their peers. An example FP showing how data were displayed to the pathologists is shown in Figure 1. A random audit of 400 specimens showed zero errors (100% correlation) with respect to

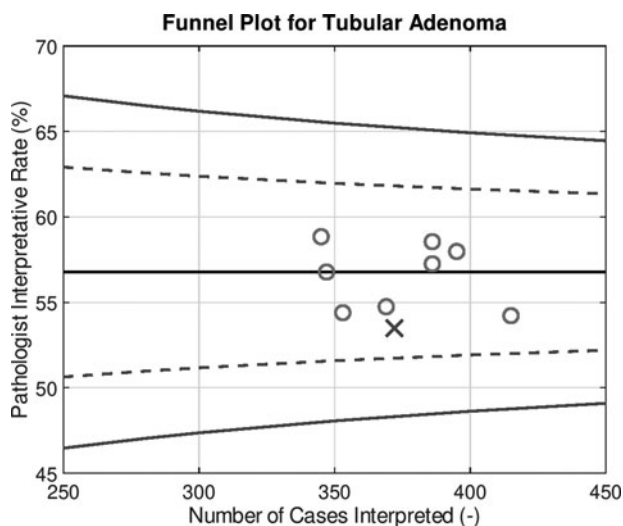


Figure 1. Funnel plot showing how data were presented to the pathologists in the study. The “X” marks the pathologist of interest. Other pathologists are marked with an “O.” The horizontal line in the center of the funnel is the group median diagnostic rate. The dashed (inner funnel) curves represent the boundaries of the 95% CI and correspond to $P < .05$. The solid (outer funnel) curves represent the boundaries of the 99.9% CI and correspond to $P < .001$. CI indicates confidence interval.

the hierarchical HFTSMA’s categorization of SSA/not SSA. An overview of the data is summarized in Tables 1 and 2. Diagnoses in the table are not mutually exclusive; 1 specimen bottle may contain 0, 1, or more polyps and consequently more than 1 diagnosis.

The raw SSA PDR (for the 9 pathologists interpreting >250 colorectal polyp specimens per year) mean/median/SD/min-max in the first and second years was 3.9%/3.7%/3.0%/0.0%-7.9% and 4.1%/3.8%/2.5%/0.8%-8.7%, respectively.

FPS/CCs for the first and second years showed 6/4 and 3/1 $P < .05/P < .001$ pathologist outliers, respectively, in relation to the group median diagnostic rate (GMDR) for SSA and 0/0 and 0/0 $P < .05/P < .001$ pathologist outliers, respectively, in relation to the GMDR for TA. The raw data, presented

Table 1. Diagnoses by Year for the Study Pathologists^a

Diagnosis	Year 1	Year 2
Tubular adenoma	1859	1692
Sessile serrated adenoma	128	124
Tubulovillous/villous adenoma	137	144
Hyperplastic polyp	792	693
Benign colorectal mucosa	420	303
Total specimens	3368	3026

^aThis table shows the frequency of common diagnoses in the 2-year study period. The categories are not mutually exclusive diagnoses (as a specimen may contain more than 1 polyp) and not all encompassing; thus, the specific diagnoses do not sum to “Total specimens.”

Table 2. Pathologists Volume Statistics^a

Statistic	Year 1	Year 2
Mean	374.22	336.22
Median	372	344
Standard deviation	23.58	38.31
Maximum	415	379
Minimum	345	265

^aThis table shows summary statistics for each of the 2 years in the study period.

via FPs, are seen in Figures 2A-2D. The data were also normalized (see SDC Appendix C, available at: <http://links.lww.com/QMH/A51>); the data in this form are presented in Figures 3A-3D.

Eight pathologists interpreted more than 250 polyps per year in both year 1 and year 2. The change in PDR for SSA is shown in Table 3; the group of 8 is ordered by the diagnostic rate in year 1 from highest to lowest. It is noteworthy that all the lowest PDR pathologists in year 1 increased their call rates.

Figures 4A and 4B show the normalized percentage of left colon and rectum polyp cases by pathologist in both years of the study; these plots are in keeping with random case assignment.

Expert-led case review

The expert-led case review was well received by the pathologists and was completed in approximately 1 hour.

In silico kappa

ISKs showed marked differences between TA and SSA. Between year 1 and year 2, minimal changes were seen in TA and a larger change was seen in SSA (see Table 4 for details). The NISK calculation showed identical trends (see Table 5 for details).

DISCUSSION

SSAs are known precursors of colorectal adenocarcinoma.¹ As such, their pathological identification is crucial to guiding patient management. The hierarchical HFTSMA generated diagnostic categories with robustness sufficient to allow insight into the practice patterns over 2 years. Using a novel custom-automated approach reviewing more than 7054 colorectal polyp specimens over 2 years, we show here that there was significant variation between pathologists in reporting SSAs compared with TAs especially in year 1, which may influence patient follow-up and, ultimately, patient outcome.

The results for TAs (used as a reference) showed no significant outliers in both years of the study. After the focused expert-led review, cohort variation in the SSA PDR remained high in relation to TA; however, diagnostic consistency for SSA appears to have increased in year 2 (number of pathologist outliers decreased, ISK increased). Targeted expert-led review appears to help calibrate the PDR and follow-up data allow

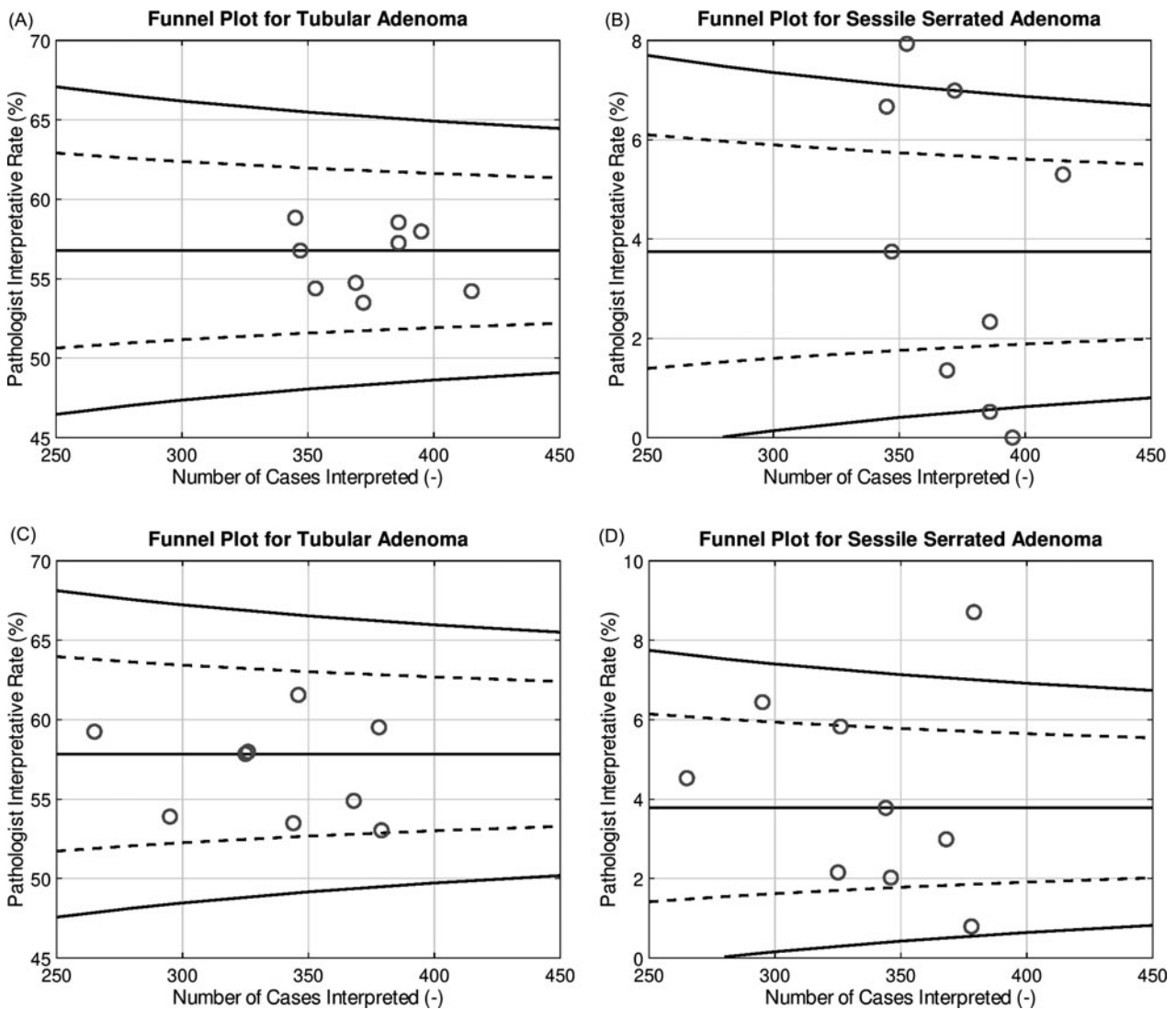


Figure 2. Funnel plots showing (A) the tubular adenoma diagnostic rate data in the first year of the study, (B) the sessile serrated adenoma diagnostic rate data in the first year of the study, (C) the tubular adenoma diagnostic rate data in the second year of the study, and (D) the sessile serrated adenoma diagnostic rate data in the second year of the study. The horizontal line in the center of the funnel is the group median diagnostic rate. The dashed (inner funnel) curves represent the boundaries of the 95% CI and correspond to $P < .05$. The solid (outer funnel) curves represent the boundaries of the 99.9% CI and correspond to $P < .001$. CI indicates confidence interval.

reassessment. Thus, ISK may represent an intuitive, useful metric and NGQ a promising approach for objectively increasing diagnostic consistency of PDR, further highlighting the importance of its use as a quality improvement strategy in pathology.

The simulated kappa (generated in silico), shown herein, improved for SSAs between year 1 ($\lambda = 0.52$) and year 2 ($\lambda = 0.62$) of the study. The simulated kappa for TA decreased slightly between year 1 ($\lambda = 0.95$) and year 2 ($\lambda = 0.93$); this change can also be seen in the FPs and CCs. As physicians are usually familiar with kappa for assessing interrater variability, this may be a useful metric that can be understood with relative ease. The variation of kappa for small numbers of specimens (<100) was considerable. This suggests that the traditional kappa in many study contexts is a

crude measure of real-world variation, as it depends strongly on the selected study set.

Some limitations in our study exist. It should be noted that the simulated pathologist interpretations assume “maximal overlap”; if one pathologist diagnoses SSA in 5% of polyps and another pathologist diagnoses it in 8% of polyps, the discordance rate will be the minimum of 3% ($8 - 5 = 3$). It is entirely possible that a lower diagnostic rate pathologist (eg, PDR = 5%) may make the diagnosis of SSA when a higher diagnostic rate pathologist (eg, PDR = 8%) does not; real-life pathologist interpretation comparisons do not necessarily have maximal overlap. Thus, the (calculated) ISK represents a best case scenario for the given set of PDRs. The ISKs (λ) (generated from the [raw] diagnostic rates) are slightly lower than the NISKs (generated from

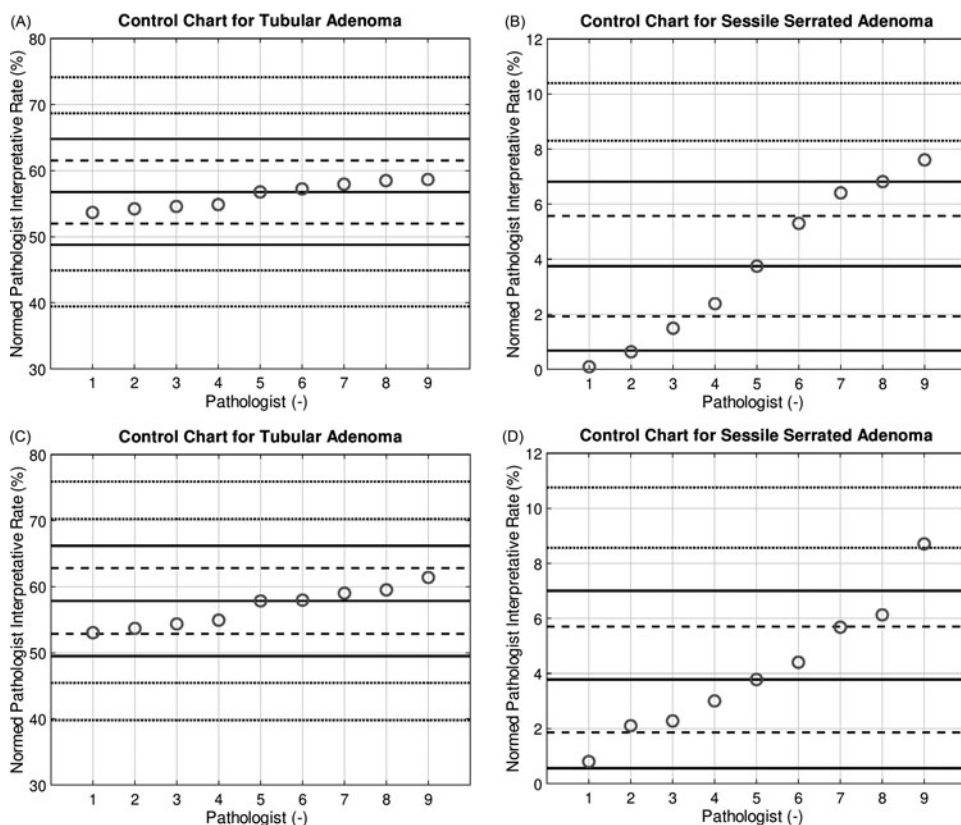


Figure 3. Control charts showing (A) the normalized tubular adenoma diagnostic rate data in the first year of the study, (B) the normalized sessile serrated adenoma diagnostic rate data in the first year of the study, (C) the normalized tubular adenoma diagnostic rate data in the second year of the study, and (D) the normalized sessile serrated adenoma diagnostic rate data in the second year of the study. The horizontal line in the center of the control chart is the group median diagnostic rate. The dashed (inner control) lines represent the boundaries of the 95% CI and correspond to $P < .05$. The solid (outer control) lines represent the boundaries of the 99.9% CI and correspond to $P < .001$. CI indicates confidence interval.

the normalized diagnostic rates); this is a consequence of how the normalization was done. Furthermore, the gastrointestinal specialist pathologist (S.A.) works at an

affiliated hospital in the same city and completed locum work at our study institution. However, she did not read more than 250 colorectal polyp specimens per year at the study site. As such, she was not included in the studied group of pathologists. A direct diagnostic rate comparison would have proved additional context to the study. In addition, the duration of the study was relatively short. Whether the change is enduring remains to be determined with further data. Outcome data by diagnostic category were not available; these would be ideal for calibrating diagnostic rate/polyp classification. The study did not make use of normalized deviations plots, a tool that allows one to identify outliers and (with the presumption of mutually exclusive categories) understand a relative overcall with associated undercall(s). Normalized deviations plots would elucidate clinical implications of relative undercalls/overall calls, as they provide the “substitute” diagnosis/diagnoses²²; we plan to do this in the future.

Table 3. Change in Diagnostic Rate With Time^a

Pathologist	$\Delta(\text{PDR})/\Delta t$
1	0.008
2	-0.012
3	0.011
4	-0.017
5	0.007
6	0.024
7	0.003
8	0.022

Abbreviations: PDR, pathologist diagnostic rate; SSA, sessile serrated adenoma.

^aThis table shows the change in the PDR from year 1 to year 2 for the 8 pathologists who read more than 250 colorectal polyp specimens per year in both years of the study. The pathologists are ordered by the SSA PDR in year 1; in this subset, pathologist 1 had the highest SSA PDR in year 1 and pathologist 8 had the lowest SSA PDR in year 1. Pathologists with negative values (eg, -0.012, -0.017) had a lower PDR in year 2 than in year 1. Pathologists with positive values (eg, 0.022, 0.003) had a higher PDR in year 2 than in year 1.

Pathologists often do not have a sense of whether they call a diagnosis frequently or infrequently as compared with their peers because they do not usually have access to their/their colleagues’ diagnostic call rates. Calculating the diagnostic rates may make one aware of and better understand differences and compare

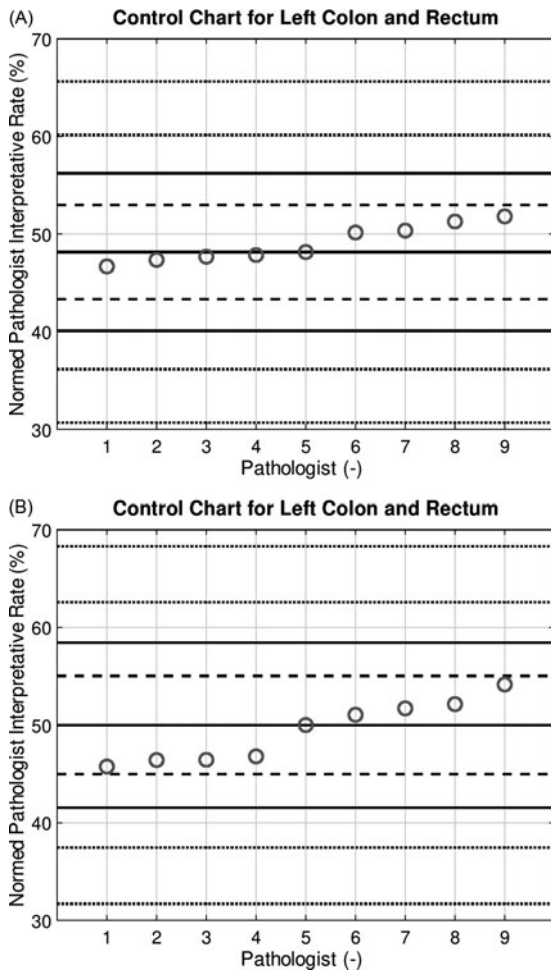


Figure 4. Control chart showing (A) the normalized left colon diagnostic rate by pathologist in the first year of the study, and (B) the normalized left colon diagnostic rate by pathologist in the second year of the study. The horizontal line in the center of the control chart is the group median diagnostic rate. The dashed (inner control) lines represent the boundaries of the 95% CI and correspond to $P < .05$. The solid (outer control) lines represent the boundaries of the 99.9% CI and correspond to $P < .001$. CI indicates confidence interval.

themselves with other institutions and the literature; thus, diagnostic rate awareness may be a useful starting point for rational discussions about the optimal pathological classification and a process to get there.

The study was based in one academic medical center where a collegial environment and an interest in

λ	Year 1	Year 2
TA λ	0.95 (0.89-0.96)	0.93 (0.86-0.94)
SSA λ	0.52 (0.43-0.58)	0.62 (0.51-0.70)

Abbreviations: SSA, sessile serrated adenoma; TA, tubular adenoma.
^aThe table shows the in silico kappa (λ) for both TA and SSA. The numbers in the brackets represent the 95% confidence interval of the variance due to rate variance (Vd2RV); this was calculated by the bootstrap method (see SDC Appendix D, available at: <http://links.lww.com/QMH/A51>, for details).

Table 5. Normalized In Silico Kappas by Year^a

NISK	Year 1 (Vd2RV)	Year 2 (Vd2RV)
TA NISK	0.95 (0.89-0.96)	0.93 (0.87-0.94)
SSA NISK	0.54 (0.45-0.60)	0.63 (0.51-0.70)

Abbreviations: NISK, normalized in silico kappa; SSA, sessile serrated adenoma; TA, tubular adenoma.
^aThis tables shows the NISK for both TA and SSA. The numbers in the brackets represent the 95% confidence interval of the variance due to rate variance (Vd2RV); this was calculated by the bootstrap method (see SDC Appendix D, available at: <http://links.lww.com/QMH/A51>, for details).

quality improvement exist among the pathologists. It is understood that work of this nature can be done within the context of a quality review without ethics approval; however, we believe that without this type of work being published and generating a dialogue, it will not happen in many environments due to lack of knowledge. Seen more broadly, it is our belief that publishing these types of studies is important to earn the trust of the public.

In conclusion, the process described appears to be useful for decreasing diagnostic disagreements. It allows one to assess a whole practice and is suited to high-volume specimens. As the process is largely automated, it is an ideal method to carry out statistical assessment. NGQ/SPC is a tool that can be used to direct quality improvement. Its wider application would enhance uniformity and lead to better patient risk stratification and likely lower costs with better outcomes. In the modern era of data mining, NGQ carries with it robust potential for future quality improvement endeavors.

REFERENCES

- Torlakovic E, Skovlund E, Snover DC, Torlakovic G, Nesland JM. Morphologic reappraisal of serrated colorectal polyps. *Am J Surg Pathol.* 2003;27(1):65-81.
- Abdeljawad K, Vemulapalli KC, Kahi CJ, Cummings OW, Snover DC, Rex DK. Sessile serrated polyp prevalence determined by a colonoscopist with a high lesion detection rate and an experienced pathologist. *Gastrointest Endosc.* 2015;81(3):517-524.
- Gurudu SR, Heigh RI, De Petris G, et al. Sessile serrated adenomas: demographic, endoscopic and pathological characteristics. *World J Gastroenterol.* 2010;16(27):3402-3405.
- Spring KJ, Zhao ZZ, Karamatic R, et al. High prevalence of sessile serrated adenomas with BRAF mutations: a prospective study of patients undergoing colonoscopy. *Gastroenterology.* 2006;131(3):1400-1407.
- Carr NJ, Mahajan H, Tan KL, Hawkins NJ, Ward RL. Serrated and non-serrated polyps of the colorectum: their prevalence in an unselected case series and correlation of BRAF mutation analysis with the diagnosis of sessile serrated adenoma. *J Clin Pathol.* 2009;62(6):516-518.
- Yang S, Farraye FA, Mack C, Posnik O, O'Brien MJ. BRAF and KRAS mutations in hyperplastic polyps and serrated adenomas of the colorectum: relationship to histology and CpG island methylation status. *Am J Surg Pathol.* 2004;28(11):1452-1459.
- Huang CS, O'Brien MJ, Yang S, Farraye FA. Hyperplastic polyps, serrated adenomas, and the serrated polyp neoplasia pathway. *Am J Gastroenterol.* 2004;99(11):2242-2255.
- Goldstein NS. Serrated pathway and APC (conventional)-type colorectal polyps: molecular-morphologic correlations, genetic pathways, and implications for classification. *Am J Clin Pathol.* 2006;125(1):146-153.

9. Winawer S, Fletcher R, Rex D, et al. Colorectal cancer screening and surveillance: clinical guidelines and rationale: update based on new evidence. *Gastroenterology*. 2003;124(2):544-560.
10. Lieberman DA, Rex DK, Winawer SJ, Giardiello FM, Johnson DA, Levin TR. Guidelines for colonoscopy surveillance after screening and polypectomy: a consensus update by the US Multi-Society Task Force on Colorectal Cancer. *Gastroenterology*. 2012;143(3):844-857.
11. Sharara N, Adam V, Crott R, Barkun AN. The costs of colonoscopy in a Canadian hospital using a microcosting approach. *Can J Gastroenterol*. 2008;22(6):565-570.
12. Mendivil J, Appierto M, Aceituno S, Comas M, Rué M. Economic evaluations of screening strategies for the early detection of colorectal cancer in the average-risk population: a systematic literature review. *PLoS One*. 2019;14(12):e0227251.
13. Khalid O, Radaideh S, Cummings OW, O'Brien MJ, Goldblum JR, Rex DK. Reinterpretation of histology of proximal colon polyps called hyperplastic in 2001. *World J Gastroenterol*. 2009;15(30):3767-3770.
14. Wong NA, Hunt LP, Novelli MR, Shepherd NA, Warren BF. Observer agreement in the diagnosis of serrated polyps of the large bowel. *Histopathology*. 2009;55(1):63-66.
15. Ensari A, Bilezikci B, Carneiro F, et al. Serrated polyps of the colon: how reproducible is their classification? *Virchows Arch*. 2012;461(5):495-504.
16. Glatz K, Pritt B, Glatz D, Hartmann A, O'Brien MJ, Blaszyk H. A multinational, Internet-based assessment of observer variability in the diagnosis of serrated colorectal polyps. *Am J Clin Pathol*. 2007;127(6):938-945.
17. Rex DK, Ahnen DJ, Baron JA, et al. Serrated lesions of the colorectum: review and recommendations from an expert panel. *Am J Gastroenterol*. 2012;107(9):1315-1330.
18. Pathology Working Group. *National Colorectal Cancer Screening Network Classification of Benign Polyps*. Toronto, ON, Canada: Canadian Partnership Against Cancer; 2011:1-28. <http://www.cdha.nshealth.ca/system/files/sites/77/documents/2pathology-working-groupphase-1-reportfinalnov-2011.pdf>. Accessed December 10, 2020.
19. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med*. 2005;24(8):1185-1202.
20. Bragg F, Cromwell DA, Edozien LC, et al. Variation in rates of caesarean section among English NHS trusts after accounting for maternal and clinical risk: cross sectional study. *BMJ*. 2010;341:c5065.
21. Mayer EK, Bottle A, Rao C, Darzi AW, Athanasiou T. Funnel plots and their emerging application in surgery. *Ann Surg*. 2009;249(3):376-383.
22. Bonert M, El-Shinnawy I, Carvalho M, et al. Next Generation Quality: assessing the physician in clinical history completeness and diagnostic interpretations using funnel plots and normalized deviations plots in 3,854 prostate biopsies. *J Pathol Inform*. 2017;8:43.
23. Bonert M, El-Shinnawy I, Rahman M, et al. Immunohistochemistry use by diagnostic category and pathologist in 4477 prostate core biopsy sets assessed at two hospitals. *Appl Immunohistochem Mol Morphol*. 2020;28(4):259-266.
24. Fisher LM, Humphries BL. Statistical Quality Control of rabbit brain thromboplastin for clinical use in the prothrombin time determination. *Am J Clin Pathol*. 1966;45(2):148-152.
25. Luttmann RJ. Next Generation Quality, part 1: gateway to clinical process excellence. *Top Health Inf Manage*. 1998;19(2):12-21.
26. Luttmann RJ. Next Generation Quality, part 2: balanced scorecards and organizational improvement. *Top Health Inf Manage*. 1998;19(2):22-29.
27. Thor J, Lundberg J, Ask J, et al. Application of statistical process control in healthcare improvement: systematic review. *Qual Saf Health Care*. 2007;16(5):387-399. doi:10.1136/qshc.2006.022194.
28. World Health Organization. Ethical issues in patient safety research. http://apps.who.int/iris/bitstream/handle/10665/85371/9789241505475_eng.pdf;jsessionid=FEB84E29E309113A478A26B6546586CD?sequence=1. Accessed April 15, 2019.
29. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257-268.