Variability in cecal intubation rate by calculation method

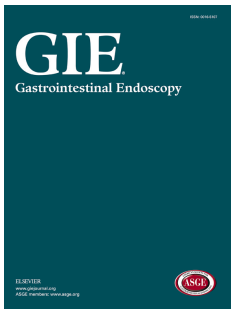
Siau, Keith; Hodson, James; Ravindran, Srivathsan; Rutter, Matthew D.; Iacucci, Marietta; Dunckley, Paul

DOI:
10.1016/j.gie.2018.12.026

Document Version
Peer reviewed version

Citation for published version (Harvard):
Siau, K, Hodson, J, Ravindran, S, Rutter, MD, Iacucci, M & Dunckley, P 2019, 'Variability in cecal intubation rate by calculation method: a call for standardization of key performance indicators in endoscopy', *Gastrointestinal Endoscopy*, vol. 89, no. 5. https://doi.org/10.1016/j.gie.2018.12.026

Accepted Manuscript



variability in cecal intubation rate by calculation method: a call for standardization of key performance indicators in endoscopy

Keith Siau, James Hodson, Srivathsan Ravindran, Matthew D. Rutter, Marietta Iacucci, Paul Dunckley

PII: S0016-5107(19)30014-8

DOI: https://doi.org/10.1016/j.gie.2018.12.026

Reference: YMGE 11394

To appear in: *Gastrointestinal Endoscopy*

Received Date: 28 October 2018

Accepted Date: 29 December 2018

Please cite this article as: Siau K, Hodson J, Ravindran S, Rutter MD, Iacucci M, Dunckley P, variability in cecal intubation rate by calculation method: a call for standardization of key performance indicators in endoscopy, *Gastrointestinal Endoscopy* (2019), doi: https://doi.org/10.1016/j.gie.2018.12.026.

VARIABILITY IN CECAL INTUBATION RATE BY CALCULATION METHOD: A CALL FOR STANDARDIZATION OF KEY PERFORMANCE INDICATORS IN ENDOSCOPY

Keith **Siau**[1,2], James **Hodson**[3], Srivathsan **Ravindran**[4], Matthew D **Rutter**[5,6], Marietta **Iacucci**[1,3], Paul **Dunckley**[2,7]

Institutions:

[1] NIHR Biomedical Research Centre, University of Birmingham, Birmingham, UK
[2] Joint Advisory Group on Gastrointestinal Endoscopy, Royal College of Physicians, London, UK
[3] Institute of Translational Medicine, University Hospital Birmingham, Birmingham, UK
[4] Wolfson Endoscopy Unit, St Mark's Hospital, Harrow, UK
[5] Department of Gastroenterology, University Hospital of North Tees, Stockton-on-Tees, UK
[6] Northern Institute for Cancer Research, Newcastle University, UK
[7] Department of Gastroenterology, Gloucestershire Hospitals NHS Foundation Trust, Gloucester, UK

Corresponding author: Dr Keith Siau; keithsiau@nhs.net

Keywords:
Caecal intubation rate, Endoscopy, Colonoscopy, Training, Metrics, Competence, Key Performance Indicators

Competing Interests:
The authors declare no conflicts of interest.

Author contributions: KS drafted the original manuscript and performed the statistical analyses. JH provided statistical insight and support. SR assisted with literature review. MDR, MI and PD provided critical review and enhanced the quality of the manuscript.

Word Count: **Abstract**: 259, **Manuscript**: 3323

ABSTRACT

Background and Aims: The cecal intubation rate (CIR) is a widely accepted key performance indicator (KPI) in colonoscopy but lacks a universal calculation method. We aimed to assess whether differences in CIR calculation methods could impact on perceived trainee outcomes.

Methods: A systematic review of CIR calculation methods was conducted on major societal guidelines (United Kingdom, European Society of Gastrointestinal Endoscopy [ESGE] and American Society for Gastrointestinal Endoscopy [ASGE]) and trainee-inclusive studies. Trainees awarded colonoscopy certification between June 2011 and 2016 were identified from the United Kingdom e-portfolio and selected as a validation cohort. For each trainee, both the crude and unassisted CIR were calculated over 50 post-certification procedures using definitions from the 3 international guidelines. The resulting CIRs, and the proportions of endoscopists failing to meet the minimum standard of CIR ≥90%, were then compared across these definitions.

Results: Across the 3 guidelines and 37 eligible studies identified, differences in CIR calculation methodology were demonstrated. These related to adjustment criteria (18 studies), and whether unassisted CIR was stipulated (18 studies). In the validation cohort of 733 trainees (36,650 procedures), the median crude CIR ranged from 96% (ESGE) to 98% (ASGE) [p<0.001], and whether unassisted CIR was specified (ESGE: 94%, ASGE: 96%, P<0.001). The proportion of trainees failing to achieve CIR ≥90% varied significantly across the different definitions, from 4.9% in the crude ASGE definition, to 18.6% in the unassisted ESGE definition (p<0.001).

Conclusions: CIR calculation methods vary between guidelines and research studies, which impact on trainee performance measures. With CIR used as an example, this study highlights the need for standardized definitions and calculations of KPIs in endoscopy.

INTRODUCTION

Colonoscopy is the criterion standard modality for investigating the lower GI tract. The ability to complete the examination and visualise all segments of the colon is a fundamental requirement of diagnostic colonoscopy. Failure to achieve completion can lead to missed pathology, additional investigations and delays in diagnoses, to the detriment of the patient and service provider. The cecal intubation rate (CIR) is an established key performance indicator (KPI), which measures an endoscopist's ability to achieve colonoscopy completion. CIR improves with training,[1,2] positively correlates with adenoma detection rate,[3] and inversely correlates with rates of postcolonoscopy colorectal cancer.[4] Consequently, the CIR features ubiquitously within international quality standards for colonoscopy and is monitored within endoscopy units.

Despite the plethora of research devoted to colonoscopy quality and performance metrics, there remains a lack of clarity on the precise definitions and calculation methods for KPIs such as CIR. KPI calculations are dependent on a number of variables (**Figure 1**), including (1) numerator, (2) denominator, (3) exclusion criteria, (4) the study window, ie, time period or number of procedures over which the KPI may be measured, (5) modifiers, eg, unassisted completion, and where applicable, (6) summary statistic, eg, averages presented either as mean, median or pooled rate (**Supplementary Table 1**). Worldwide, KPI calculations vary between guidelines. In addition, renewed attention to colonoscopy quality assurance has led to modern-day dependence on endoscopy reporting systems (ERS) to automate the processing of unit-level procedural data and KPI outputs.[5-7] Despite efforts to standardise electronic data collection,[8] the heterogenous nature of ERS platforms may also lead to variations in KPI calculation methodology. Clearly defined and accurate measurements of colonoscopy KPIs are therefore required to ensure consistency of KPI outcome reporting.

Within the realm of colonoscopy training, CIR is used as a performance metric to monitor a trainee's progress and inform decisions on certification/credentialing, which enables trainees to practice independently.[1,2,9] Calculations which overestimate CIR may enable trainees to begin service provision before developing full competence, whereas underestimation may unfairly penalise trainees, stifle progression toward certification and reduce the availability of endoscopists to provide service. Thus, ensuring the validity of CIR calculations in the training cohort is paramount.

This study had the following aims and objectives:

1) Compare CIR calculation methodology between international guidelines.
2) Perform a systematic review to assess for heterogeneity in CIR calculation methodology within the colonoscopy training literature.
3) Assess whether differences in CIR calculation methods could affect endoscopist performance measures.

METHODS

Study Design

This study consisted of 2 components: (1) a systematic review of CIR calculation methods relevant to trainee endoscopists, and (2) application of CIR calculations to a United Kingdom (UK) training cohort using the national training e-portfolio, in order to evaluate the impact of different CIR calculations on trainee outcomes.

Systematic Review

A systematic review of CIR calculation methods within the training literature was performed in accordance with PRISMA recommendations.[10] The literature search was conducted on EMBASE, and PubMed using a combination of the following terms: "Cecal /cecal intubation rate," MESH heading "colonoscopy" AND "train* OR fellow." Studies were eligible if any trainees/fellows were included within the cohort. Exclusion criteria comprised the following: non-English main text, meta-analyses, non-human studies, or where CIR was not a study outcome. Full-text assessments were performed in order to determine the method of CIR calculation.

Data were systematically extracted (by 2 independent authors: K.S. and V.R.) into the following fields:

a) Author
b) Year of publication
c) Study design
d) Verification of cecal intubation, ie, self-reported (based on electronic records documenting visualization of landmarks), direct observation, or review of photodocumentation.
e) Exclusions
f) Unassisted CIR (specification of physically unassisted/independent cecal intubation)
g) Study window, ie, whether CIR was studied by blocks of procedures or provided across an entire study period.
h) Summary statistic

Trainee Validation Cohort

We assessed the impact of CIR calculations on competency determination by performing a validation study on a national UK-wide cohort of colonoscopy trainees. The Joint Advisory Group on Gastrointestinal Endoscopy (JAG) is the advisory body tasked with quality assuring endoscopy training and service within the UK[11] JAG colonoscopy certification is required for independent practice and is awarded in 2 stages: provisional and full, the criteria for which have been previously described, but include CIR as a marker of competence.[2] Provisional colonoscopy certification (PCC) enables endoscopists to perform diagnostic colonoscopy without immediate supervision. This study was a *post hoc* analysis of the PCC dataset,[2] which involved a national trainee-maintained database of endoscopy procedures: the JAG Endoscopy Training System (JETS) e-portfolio,[12] a web-based platform for the recording and verification of endoscopy experience and for awarding certification. Colonoscopy procedures of trainees who were awarded certification between 2011 and 2016 were identified; trainee entries in the 50 procedures after certification were selected as this corresponds to the period of potentially newly independent endoscopy practice, where practitioners remain under a period of performance monitoring until full certification is granted. For each procedure, variables collected within the e-portfolio were extracted, including: procedure extent, quality of bowel preparation (using the modified Aronchick scale),[13] emergency versus nonemergency cases, need for physical assistance during intubation, diagnoses, and reasons for failed completion. Physical assistance was defined as the need for another endoscopist to take-over the endoscope at

any point of insertion to the cecum. In line with the Aronchick scale, poor/inadequate bowel preparation was defined as <90% mucosa seen with a mixture of semisolid and solid stool that could not be suctioned or washed. Colonoscopy cases were based on unselected cases; however, this newly certified cohort were exempt from performing Bowel Cancer Screening procedures and complex therapeutic colonoscopy.

Outcomes

The primary outcome studied was the CIR as defined by each of the guidelines of interest, ie, crude CIR. As the intention of CIR is to depict an endoscopist's technical skills, the unassisted CIR was also presented. Within the validation cohort, CIR was studied as a continuous variable over 50 consecutive post-certification procedures; the crude CIR was inferred from the overall procedural extent, and unassisted CIR from the most proximal trainee-specific extent. Procedures that required physical assistance from another endoscopist to achieve cecal intubation were therefore included in the crude CIR calculation, but not for unassisted CIR.

The secondary outcome was the percentage of trainees who surpassed minimum competency standards based on each CIR calculation, using the previously presented outcome of drop in performance (DIP),[2] defined as CIR <90% over the first 50 procedures post-certification. DIP rates were presented for both crude and unassisted CIR.

Statistical Analyses

Initially, the CIR was calculated for each trainee using each of the international guidelines identified. The average CIRs were then compared across the guidelines using the Friedman test, as the data followed a skewed distribution, which was followed by pairwise post-hoc comparisons when the overall effect was significant. For each guideline, comparisons were then performed between the crude and unassisted definitions of CIR using the Wilcoxon tests. Similar analyses were then performed for the outcome of DIP, with McNemar's test used to compare the crude and unassisted definitions.

Continuous variables were summarized using medians and interquartile ranges (IQRs). All analyses were performed using IBM SPSS 22 (IBM Corp. Armonk, NY), with $p<0.05$ deemed to be indicative of statistical significance throughout.

RESULTS

Summary of CIR calculations in International Guidelines

Currently, 4 main guideline documents have set performance standards in colonoscopy of relevance to trainees. The original JETS certification criteria require the attainment of an unassisted and unadjusted CIR of 90%.[2,14] The UK quality standards recommendations, released in 2016,[15] also base CIR calculations on an unadjusted rate. More recently, the European Society of Gastrointestinal Endoscopy (ESGE) performance standards recommend a CIR of 90%.[16] This was the first guideline to explicitly define the numerator and denominator for the CIR calculation by recommending the exclusion of emergency procedures, which tend to be associated with poor bowel preparation and incomplete colonoscopy. Additionally, procedures whereby colonoscopy had reached therapeutic intent, eg, polypectomy in the transverse colon after a previously completed procedure, could be excluded from the calculation. The American Society for Gastrointestinal Endoscopy (ASGE) guideline makes allowances for inadequate bowel preparation or severe colitis,[17] enabling the omission of procedures with failed completion from the CIR denominator, provided that there is supporting photodocumentation. All societies stipulate a minimum standard CIR of 90%, with the non-training guidelines recommending an aspirational standard of 95%, particularly within the bowel cancer screening setting. A summary of the guidelines and calculation methods is provided in **Table 1**.

Literature Review

The search strategy yielded 198 results (**Supplementary Figure 1**), of which 37 studies met eligibility criteria and were included in the systematic review (**Table 2**). [1,2,5,9,18-50] From the studies identified, cecal intubation was verified using self-reported methods (N=21), direct observation (N=12) and photodocumentation (N=3). Eighteen studies (49%) applied an adjusted CIR, with 3 studies excluding failed procedures from the denominator. The main exclusion criteria comprised: previous colonic surgery (N=10), colitis (N=9); colonic obstruction, eg, stricture (N=9); emergency presentation, eg, lower gastrointestinal bleeding (N=9); inadequate bowel preparation (N=7); therapeutic intent, eg, planned polyp resection (N=7). Unassisted CIR was clearly defined in 18 studies (49%) and implied as unassisted (based on CIR results) in an additional 3 studies. For the averaging method, the majority relied on the pooled rate (N=24, 65%), with 14 studies (39%) involving a study window of ≥6 months.

Trainee Validation Cohort

For each of the 733 trainees included in the analysis, outcomes of the 50 procedures after JAG certification were collated, giving a total of 36,650 procedures for analysis. The trainee characteristics of this validation cohort, and reasons for failed cecal intubation, have been previously described.[2] Because data were only recorded for N=50 procedures per trainee, we used this as the basis of all of the calculations of CIR, rather than the ≥100 consecutive procedures specified by the ESGE guideline. As the validation cohort were effectively independent practitioners, performance was compared at trainee-level between the guidelines relevant to independent practitioners, ie, UK, ESGE, ASGE.

The UK guidelines made no exclusions, meaning that all 50 procedures were used to calculate the CIR for each trainee. For ESGE guidelines, emergency procedures (N=62, 0.2%) were excluded; there were no procedures without intent to reach the cecum in this cohort. For the ASGE guidelines, a total of 674 (1.8%) procedures were excluded due to poor bowel preparation or severe colitis.

CIR by guideline

The UK and ESGE guidelines returned similar crude CIRs, with both having a median of 96% and IQR of 92-98% (post-hoc test p=1.000). However, the larger number of procedures excluded by the ASGE guidelines resulted in a significantly higher CIR, with median of 98% (IQR: 96-100%, post-hoc test p<0.001 vs UK and ESGE; **Table 3**). The data were then further interrogated, to identify differences between the guidelines on a trainee level (**Figure 2**). The UK and ESGE guidelines returned identical crude CIRs for 97% (N=712) of trainees. The CIR was higher on the UK guideline in 2% (N=17), with remainder (0.1%, N=4) being higher on the ESGE guideline. When comparing the UK and ASGE guidelines, only 47% (N=341) trainees had identical crude CIRs. No trainees had higher CIRs on the UK guideline, with 44% (N=321) of CIRs being up to 5 percentage points higher using the ASGE measurement, and 10% (N=71) differing by a greater margin. The largest observed difference between these guidelines was 14 percentage points, in a trainee achieving a crude CIR of 84% (42/50) on the UK guideline, compared with 98% (42/43) on the ASGE guideline. For this trainee, all 7 incomplete procedures were attributed to inadequate bowel preparation.

Impact of Adjusting for Physical Assistance (Unassisted CIR)

Comparisons between the crude and unassisted definitions of CIR found the former to be significantly higher for all 3 guidelines, with medians of 96% vs. 94% (p<0.001) for both the UK and ESGE guidelines, and 98% versus 96% (p<0.001) for the ASGE guideline (**Table 3**). Calculations using the unassisted definition resulted in 40% (N=293) having a lower CIR than if the crude definition had been used. As a result, the proportion of endoscopists with DIP was higher when the unassisted definition was used for each guideline (all p<0.001), with a 2-fold to 3-fold increase in the proportion of patients failing to meet the minimum standard, relative to the crude CIR (**Figure *3***).

Impact of Averaging Method

None of the guidelines clearly reported the summary statistic that should be used for averaging the CIR across a cohort of endoscopists. There are 3 common statistics that could be used, namely a median, mean or pooled rate, the impact of which are summarized with examples in **Supplementary Table 1**. Because the distribution of CIR is generally skewed (**Figure 4**), reporting a mean is likely to be influenced by outliers with low CIR in the "tail" of the distribution. If the number of procedures included in the calculation of CIR is the same for each endoscopist, then the pooled rate and mean will return identical results. However, if the definition of CIR only requires a minimum number of

procedures (eg, ≥100 in the ESGE definition), then those with a greater number of procedures will have a greater influence on the pooled rate.

In our study, for the crude CIR by the UK guideline, both the mean and pooled rate approaches yield a CIR of 94.8%, because all trainees are contributing the same number of procedures (N=50). However, this is lower than the median value of 96%, on account of the skew in the distribution.

DISCUSSION

In pursuit of quality in endoscopy, attention has turned toward the performance of individual endoscopists and the use of KPIs as a proxy measure. This study demonstrates that significant heterogeneity exists in the calculation of the CIR KPI by reporting method, both within research studies but also within major international guidelines. These have a subsequent impact on the interpretation of performance outcomes. For instance, endoscopists considered competent when measured according to ASGE definitions may be categorized as an underperformer using UK/European definitions. Differences arise not only due to adjusted versus unadjusted calculation methodology, but also according to how these adjustments are made. The ESGE adjustment excludes procedures from the numerator and denominator, whereas the ASGE adjustment excludes failed procedures due to specific criteria, therefore only affecting the denominator.

To ensure validity, KPIs must be conceptually representative of what they were designed to measure. KPIs may serve as surrogates for both quality of patient care and quality of the performance of an individual endoscopist. Although the measurement of colonoscopy completion (crude CIR) is relevant for patient care, in order to evaluate an endoscopist's ability to achieve independent completion, the unassisted CIR is required to ensure validity. From our systematic review, unassisted CIR was defined in 49% of trainee-inclusive studies, with the majority of studies (57%) deducing outcomes from self-reported measures involving electronic records. Our study was centred on the JETS e-portfolio, which enables trainee extent, overall procedural extent and the need for physical assistance to be recorded, thereby enabling the measurement of both crude and unassisted CIR. At present, not all ERS platforms have the capability of recording whether physical assistance has been provided to an independent endoscopist.[7] An independent endoscopist requesting physical assistance to achieve cecal intubation would be marked on the ERS as achieving at least a cecal extent. Consequently, upon interrogation of the ERS, the procedure would be interpreted as achieving cecal intubation, which could mask underperformance. Training programmes basing certification-related decisions on KPIs should ensure that these are founded on unassisted rates, otherwise this risks measuring the performance of the assisting endoscopist. This concept extends to other KPIs appraised by training programs such as the polyp/adenoma detection rate. Thus, in the modern era in which increasing reliance is placed on ERS platforms to autogenerate KPI calculations and contribute to mandatory quality assurance audits, it is pivotal for these systems to capture specific roles within a procedure and whether physical assistance by

another endoscopist has been enlisted. This is required to ensure validity of the KPI and for governance purposes, ie, recording the specific involvement of another endoscopist.

The lack of a universal CIR calculation method is evident. The ESGE working group reported that "no conclusions can be drawn about the best definition of complete colonoscopy examination because no evidence was found." [51] Similar dilemmas may be generalisable to other KPIs such as the adenoma/polyp detection rate,[52] where differences in calculation methods may also affect the outcome, eg, inclusion of colorectal cancer or sessile serrated lesions within the numerator; exclusion of colonoscopy scheduled for intended polypectomy, incomplete procedures, previous bowel resection, emergency cases or those with inadequate bowel preparation; and stratification by patient age and procedural indication. Co-ordinated international efforts involving multisociety consensus-based processes are hence necessary to harmonise definitions and methods for CIR and other KPI calculations, including each variable featured within the KPI equation. Furthermore, these guideline development groups should consider the unintended consequences of KPI measurements. Applying an unadjusted CIR threshold may lead to endoscopists attempting cecal intubation "at all costs," eg, striving for completion despite poor quality bowel preparation, and a disinclination to perform colonoscopy in settings associated with higher risk of completion failure,[35] which can be a disservice to patients. Conversely, excluding procedures with poor/inadequate bowel preparation from CIR calculations may lead to incomplete procedures being falsely attributed to poor bowel preparation. Indeed, manipulation of self-reported systems data, ie, gamification, is another potential issue. Other examples of CIR gamification, and potential solutions to overcome this, include:

1) Exaggeration of procedural extent, ie, indicating procedural completion when this has not been the case. In response, the major societies now advocate the photodocumentation of completion landmarks, although this remains inconsistently practised. The ASGE is the only society which includes photodocumentation within the CIR numerator.
2) The inappropriate conversion of incomplete colonoscopy procedures to flexible sigmoidoscopy,[22,53] which could artificially elevate CIR. This is challenging to automate and requires manual audit.
3) Self-maintained trainee portfolios risk selection bias, whereby failed procedures could be omitted to favour certification. The UK National Endoscopy Database plans to mitigate this by autopopulating training procedures directly into the JETS e-portfolio.[11]

Other limitations merit discussion. This was a cohort of newly independent colonoscopists who were typically shielded from emergency cases or specific therapeutic indications, which may have led to a lack of a significant difference between UK and ESGE calculations. Although it is recognized that CIR may be affected by a myriad of factors such as quality of training, type of endoscopy unit, endoscopist experience, sedation, and available technology, these were not reported within our validation study, as it was designed to investigate CIR measurement (ie, variables within the CIR formula) rather than CIR performance. Next, the impact of varying study windows was not assessed. For trainee endoscopists, it has previously been demonstrated that progression, as measured using unassisted CIR, occurs in a near linear fashion over lifetime procedure count.[2,35] Thus, for the

trainee cohort, measuring KPIs over blocks of 100 procedures, as per ESGE guidelines, may not be representative of current performance. The JETS criteria calculate trainee-specific KPIs using data from the preceding 3-month study window, over which at least 15 procedures are required. Although this may be more feasible, this may lead to wider confidence intervals in the CIR. As such, study windows adapted for the training setting may need to be separately defined in international guidelines. Finally, possible data manipulation borne from performance monitoring, eg, verification of photodocumentation, was not possible within the functionality of the JETS e-portfolio.

The advent of National Endoscopy Database initiatives are well placed to provide standardization of calculation measures,[54] however, refinements to the CIR calculations may be required to account for physical assistance (+/-gaming) in order to provide validity to the concept of CIR. For research and recommendations centred on KPI outcomes, greater transparency in the methodology of KPI calculations is required to ensure consistent reporting and interpretation.[55]

In conclusion, we have shown that differences in KPI calculations impact on perceived endoscopist outcomes and their inferences of competence. Using CIR as an example, we highlight in this study the need for standardized definitions and calculations of KPIs in endoscopy.

TABLES

KPI Component	JAG [14]	UK [15]	ESGE [16]	ASGE [17]
Definition	Rate of unassisted intubation to the cecum, terminal ileum or ileo-colonic anastomosis.	Percentage of colonoscopies reaching and visualizing the whole cecum and landmarks		
Standard	≥90%	Minimum: ≥90% Aspirational: ≥95%		
Numerator	Cecal completion (unassisted)	Cecal completion (ideally with photodocumentation)		
Denominator	All colonoscopy procedures			
Exclusion	None		Emergency procedures, therapy without intention to reach cecum.	Incomplete procedures due to poor bowel prep or severe colitis
Time window	Preceding 3 months	*Not stated*	≥100 consecutive procedures	*Not stated*
Averaging method	*Not stated – mean inferred*			

Table 1: Variation in cecal intubation rate calculations by international guideline. *KPI: Key performance indicator; JAG: Joint Advisory Group on Gastrointestinal Endoscopy; ESGE: European Society of Gastrointestinal Endoscopy; ASGE: American Society for Gastrointestinal Endoscopy.*

Study	Year	Design	Unassisted CIR	Verification of cecal intubation	Exclusion(s)	Study Window	Summary statistic
Valori [18]	2018	Prospective cohort	N	Self-reported	Obstructing lesions	Study period (1 year)	Pooled rate; % achieving composite endpoint
Siau [2]	2018	Prospective cohort	Y	Self-reported	None	10 procs	Mean (moving average); median
Tang [19]	2018	RCT	Y	Direct observation	Age <18 or >90 years, pregnancy, colonic resection, diverticulitis (<1 month), colonic obstruction, severe LGIB, referral for EMR, unsedated procedure	Study period (3 months)	Pooled rate
El-Halabi [20]	2018	Retrospective observational	N	Self-reported	Non-screening, nonsurveillance procedures, therapeutic intent, colonic resection, failed procedures due to inadequate prep	Study period (9 months)	Mean
Singh [21]	2017	Retrospective observational	N	Self-reported	None	Study period (4 years)	Pooled rate
Beg [22]	2017	Retrospective observational	N	Self-reported	None	Study period (12 months)	Range, % with CIR<90%
Pace [23]	2016	Retrospective cohort	N	Self-reported	None	Study period (6 months)	Pooled rate
McCarthy [24]	2016	Retrospective cohort	Y	Self-reported	None	50 procedures	Mean
Patwardhan [25]	2016	Prospective observational	Y	Self-reported	None	Study period (5 months)	Pooled rate
Klare [26]	2015	Retrospective cohort	N	Self-reported	Poor bowel preparation, “inappropriate investigations,” stenosis, therapeutic intention	N/A	N/A
Hui [27]	2015	RCT	Y	Self-reported	Previous colorectal surgery, IBD colonic adenoma or CRC, pregnancy	Study period (15 months)	Pooled rate
Walsh [28]	2015	Prospective observational	N	N/A	None	20 procedures	Pooled rate
Koch [29]	2015	Prospective cohort	Y (<20 mins)	Direct observation	Previous colonic resection, previous incomplete colonoscopy	2 procedures	Mean
McClellan [30]	2015	Retrospective observational	N	Self-reported	IBD, missing information on “cecal attainment”	Study period (3 years)	Pooled rate

Williams [31]	2015	Retrospective observational	N	Self-reported	Emergency or non-screening procedures, <50 yrs, failed procedures due to inadequate bowel preparation	6 months	Pooled rate
Kim [32]	2014	Randomized study	Y (<20 mins)	Photodocumentation	Emergency, colonic obstruction, intended therapeutic colonoscopy, history of abdominopelvic surgery, IBD surveillance	30 procedures	Pooled rate
McIntosh [33]	2014	RCT	Y	Direct observation	None	5 procedures	Pooled rate
Nemoto [34]	2014	Prospective observational	N	Direct observation	None	Study period (15 months)	Pooled rate / Mean
Ward [1]	2014	Prospective observational	Y	Self-reported	None	20 procedures	Mean (moving average)
Park [35]	2013	Prospective observational	Y (<15 mins)	Photodocumentation	Emergency, intended therapeutic colonoscopy, colonic obstruction, colorectal surgery	50 procedures	Pooled rate
Koch [9]	2012	Prospective cohort	Y	Self-reported	None	20 procedures	N/A
Luo [36]	2012	RCT	N	Direct observation	Previous colonoscopy, colorectal surgery, known stricture/tumour, severe colitis, ischaemic colitis, acute lower GI bleeding	Study period (10 months)	Pooled rate
Manta [37]	2012	Randomized study	N	Direct observation	Age <18 and >90 years, pregnancy, previous colorectal surgery, obstructing tumour, extrinsic colon compressions not allowing colonoscope transit	Study period (1 year)	Pooled rate / Mean
Selvasekar [38]	2012	Prospective, observational	Y (implied) <35 mins	Self-reported	Cancer surveillance, IBD, gastrointestinal haemorrhage	Undefined moving average	Pooled rate, % achieving outcome
Portocarrero [39]	2012	Prospective observational	Y (implied)	Direct observation	None	17 patients	N/A (N=1)
Park [40]	2012	Randomized study	Y	Direct observation	Colonic resection, fulminant colitis, severe LGIB, poor/inadequate bowel preparation	10 procedures	Pooled rate
Van Putten [41]	2012	Prospective observational	Y	Self-reported	None	100 procedures	Median
Kaltenbach [42]	2011	Prospective cohort	Y (implied)	Direct observation	None	Study period (6 weeks)	Pooled rate
Spier [43]	2010	Prospective observational	Y	Self-reported	None	Study period (2 months)	Pooled rate

Tee [44]	2010	RCT	Y	Direct observation	Colonic resection, pregnancy, severe colitis, ischaemic colitis, referral for EMR, LGIB	200 procedures	Pooled rate
Koornstra [45]	2009	Prospective observational	Y	Direct observation	None	25 procedures	Mean (moving average)
Leung [46]	2008	Retrospective observational	N	N/A	Inadequate bowel preparation, obstructing lesion	Study period (34 months)	Pooled rate
Kondo [47]	2007	RCT	Y (<15 mins)	Direct observation	None	-	N/A
Aslinia [48]	2006	Retrospective cohort	N	Self-reporting	Unadjusted CIR: None; Adjusted CIR: Poor prep or severe colitis; Circumstance adjusted CIR: Decision not to attempt cecal intubation, e.g. therapeutic intent	Study period (6 years)	Pooled rate
Bowles [49]	2004	Prospective observational	N	Self-reported	None	Study period (4 months)	Pooled rate
Cotton [5]	2003	Observational	N	Self-reported	None	Study period (5 years)	Median
Thomas-Gibson [50]	2002	Prospective observational	Y	Self-reported + adjustment for photodocumentation	Failed completion due to strictures or inadequate bowel preparation	Study period (2 months)	Pooled rate

Table 2: Summary of trainee-enrolled studies evaluating the outcome of cecal intubation rate (CIR). LGIB: lower gastrointestinal bleeding, EMR: Endoscopic mucosal resection, IBD: Inflammatory bowel disease, CRC: Colorectal carcinoma, N/A: data not available.

	Guideline			p-value (by guideline)
	UK	***ESGE***	***ASGE***	
Excluded Procedures	-	62 (0.2%)	674 (1.8%)	**-**
Crude CIR	96% (92% - 98%)	96% (92% - 98%)	98% (96% - 100%)	**<0.001**
Unassisted CIR	94% (90% - 96%)	94% (90% - 96%)	96% (92% - 98%)	**<0.001**
***P* value (crude vs unassisted)**	**<0.001**	**<0.001**	**<0.001**	**-**

Table 3: Variation in cecal intubation rates (CIR) international guideline. *Data are reported as N (%), or as median (IQR), as applicable. p-values for comparisons between guidelines are from the Friedman tests, and comparisons between crude and unassisted CIR are from the Wilcoxon test, with bold values significant at p<0.05. ESGE: European Society of Gastrointestinal Endoscopy; ASGE: American Society for Gastrointestinal Endoscopy.*

FIGURES

Figure 1: Generic formula for key performance indicator calculations, with cecal intubation rate (CIR) as an example. *ESGE: European Society of Gastrointestinal Endoscopy; ASGE: American Society for Gastrointestinal Endoscopy.*

Figure 2: Paired differences in crude cecal intubation rate (CIR) between guidelines. *Unlabelled bars each consist of <3% of the cohort; ESGE: European Society of Gastrointestinal Endoscopy; ASGE: American Society for Gastrointestinal Endoscopy; pp: percentage point*

Figure 3: Variation in the percentage of endoscopists with drop in performance (DIP) according to UK, European Society of Gastrointestinal Endoscopy (ESGE) and American Society for Gastrointestinal Endoscopy (ASGE) calculation methods. *Drop in performance (DIP) was defined by a CIR of <90%. Comparisons between the crude and unassisted CIR definitions were performed using the McNemar test, and were significant for all 3 guidelines (all p<0.001)*

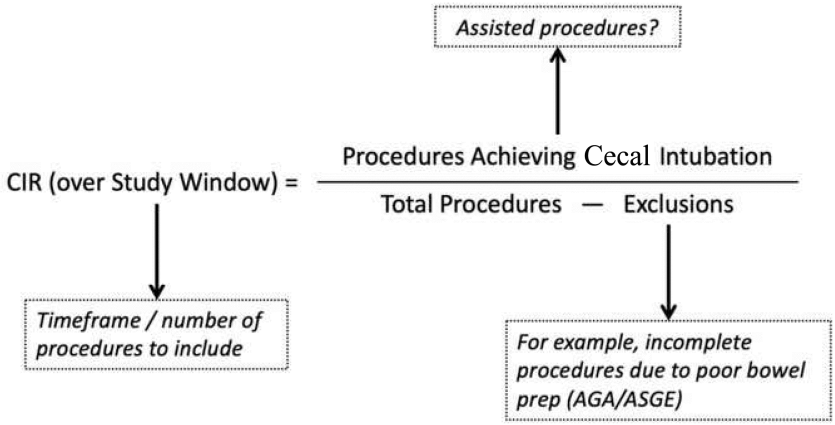
Figure 4: Histogram of crude cecal intubation rate (CIR) based on the UK guideline.

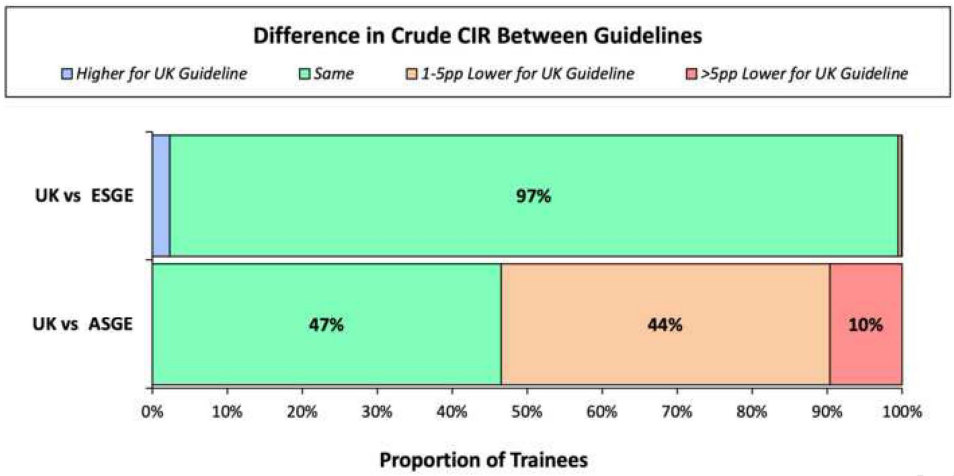
REFERENCES

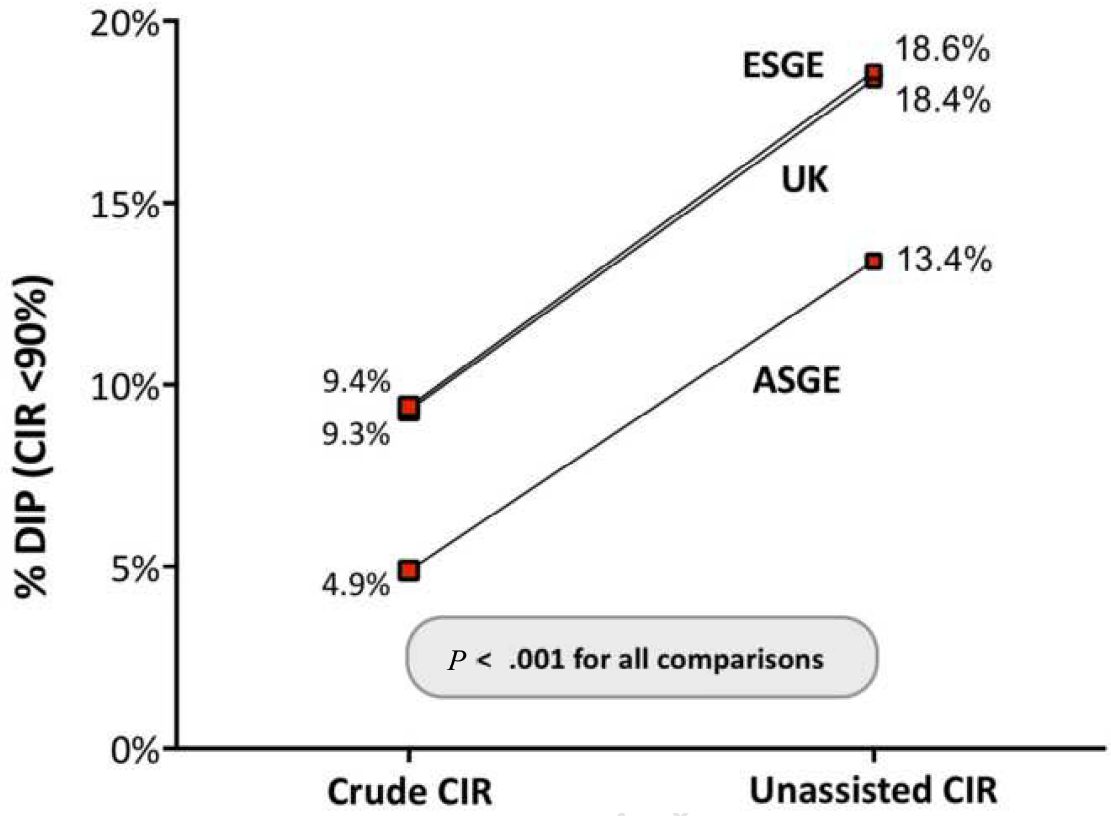
1. Ward ST, Mohammed MA, Walt R, et al. An analysis of the learning curve to achieve competency at colonoscopy using the JETS database. Gut 2014.
2. Siau K, Hodson J, Valori RM, et al. Performance indicators in colonoscopy after certification for independent practice: outcomes and predictors of competence. Gastrointest Endosc 2018 Published Online First: 2018/08/05.
3. Lee TJ, Rutter MD, Blanks RG, et al. Colonoscopy quality measures: experience from the NHS Bowel Cancer Screening Programme. Gut 2012;61:1050-7.
4. Baxter NN, Sutradhar R, Forbes SS, et al. Analysis of administrative data finds endoscopist quality measures associated with postcolonoscopy colorectal cancer. Gastroenterology 2011;140:65-72.
5. Cotton PB, Connor P, McGee D, et al. Colonoscopy: practice variation among 69 hospital-based endoscopists. Gastrointest Endosc 2003;57:352-7.
6. Manfredi MA, Chauhan SS, Enestvedt BK, et al. Endoscopic electronic medical record systems. Gastrointest Endosc 2016;83:29-36.
7. Bretthauer M, Aabakken L, Dekker E, et al. Reporting systems in gastrointestinal endoscopy: Requirements and standards facilitating quality improvement: European Society of Gastrointestinal Endoscopy position statement. United European Gastroenterology Journal 2016;4:172-76.
8. Aabakken L, Rembacken B, LeMoine O, et al. Minimal standard terminology for gastrointestinal endoscopy - MST 3.0. Endoscopy 2009;41:727-8.
9. Koch AD, Haringsma J, Schoon EJ, et al. Competence measurement during colonoscopy training: the use of self-assessment of performance measures. Am J Gastroenterol 2012;107:971-5.
10. Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. 2009;339.

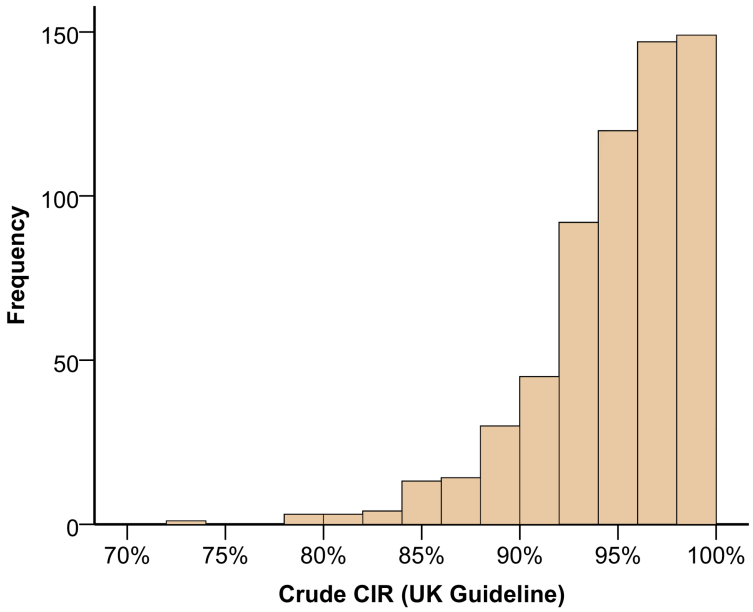
11. Siau K, Green JT, Hawkes ND, et al. Impact of the Joint Advisory Group on Gastrointestinal Endoscopy (JAG) on endoscopy services in the UK and beyond. Frontline Gastroenterology 2018.
12. Mehta T, Dowler K, McKaig BC, et al. Development and roll out of the JETS e-portfolio: a web based electronic portfolio for endoscopists. Frontline Gastroenterology 2011;2:35.
13. Aronchick C, Lipshutz W, Wright S, et al. Validation of an instrument to assess colon cleansing. 1999;94:2667.
14. Joint Advisory Group on Gastrointestinal Endoscopy. JAG trainee certification process: colonoscopy (provisional and full). 2016 [cited 27th November 2018]; Available from: https://www.thejag.org.uk/Downloads/JAG%20Certification%20for%20trainees/Colonoscopy%20application%20criteria%20and%20process.pdf.
15. Rees C, Thomas Gibson S, Rutter M, et al. UK key performance indicators and quality assurance standards for colonoscopy. Gut 2016;65:1923-29.
16. Kaminski MF, Thomas-Gibson S, Bugajski M, et al. Performance measures for lower gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) Quality Improvement Initiative. Endoscopy 2017;49:378-97.
17. Rex DK, Schoenfeld PS, Cohen J, et al. Quality indicators for colonoscopy. Gastrointestinal Endoscopy 2015;81:31-53.
18. Valori RM, Damery S, Gavin DR, et al. A new composite measure of colonoscopy: the Performance Indicator of Colonic Intubation (PICI). Endoscopy 2018;50:40-51.
19. Tang Z, Zhang DS, Thrift AP, et al. Impact of cap-assisted colonoscopy on the learning curve and quality in colonoscopy: a randomized controlled trial. Gastrointest Endosc 2018;87:723-32.e3.
20. El-Halabi MM, Barrett PR, Martinez Mateo M, et al. Should We Measure Adenoma Detection Rate for Gastroenterology Fellows in Training? Gastroenterology Res 2018;11:290-94.
21. Singh HK, Withers GD, Ee LC. Quality indicators in pediatric colonoscopy: an Australian tertiary center experience. Scandinavian Journal of Gastroenterology 2017;52:1453-56.
22. Beg S, Sansone S, Manguso F, et al. The Conversion of Planned Colonoscopy to Sigmoidoscopy and the Effect of this Practice on the Measurement of Quality Indicators. Am J Gastroenterol 2017;112(:1545-52.
23. Pace D, Borgaonkar M, Hickey N, et al. Does the hands-on, technical training of residents in colonoscopy affect quality outcomes? Surg Endosc 2016;30:1352-5.
24. McCarthy ST, Jorgensen J, Elta GH, et al. Early Splenic Flexure Intubation Competency Predicts Early Cecal Intubation Competency in Gastroenterology Fellows. Dig Dis Sci 2016;61(:3155-60.
25. Patwardhan VR, Feuerstein JD, Sengupta N, et al. Fellowship Colonoscopy Training and Preparedness for Independent Gastroenterology Practice. J Clin Gastroenterol 2016;50:45-51.
26. Klare P, Ascher S, Wagenpfeil S, et al. Trainee colonoscopists fulfil quality standards for the detection of adenomatous polyps. BMC Medical Education 2015;15:26.
27. Hui AJ, Lau JY, Lam PPY, et al. Comparison of colonoscopic performance between medical and nurse endoscopists: a non-inferiority randomised controlled study in Asia. 2015;64:1058-62.
28. Walsh CM, Ling SC, Khanna N, et al. Gastrointestinal Endoscopy Competency Assessment Tool: reliability and validity evidence. Gastrointest Endosc 2015;81:1417-24.e2.
29. Koch AD, Ekkelenkamp VE, Haringsma J, et al. Simulated colonoscopy training leads to improved performance during patient-based assessment. Gastrointest Endosc 2015;81:630-6.
30. McClellan DA, Ojinnaka CO, Pope R, et al. Expanding Access to Colorectal Cancer Screening: Benchmarking Quality Indicators in a Primary Care Colonoscopy Program. J Am Board Fam Med 2015;28:713-21.
31. Williams MR, Crossett JR, Cleveland EM, et al. Equivalence in colonoscopy results between gastroenterologists and general surgery residents following an endoscopy simulation curriculum. J Surg Educ 2015;72:654-7.
32. Kim YG, Kim K-J, Yang D-H, et al. Intermediate-length colonoscope needs more training duration than long-length colonoscope. Scandinavian Journal of Gastroenterology 2014;49:1007-13.
33. McIntosh KS, Gregor JC, Khanna NV. Computer-based virtual reality colonoscopy simulation improves patient-based colonoscopy performance. Canadian Journal of Gastroenterology & Hepatology 2014;28:203-06.
34. Nemoto D, Isohata N, Utano K, et al. Double-balloon colonoscopy carried out by a trainee after incomplete conventional colonoscopy. Dig Endosc 2014;26:392-5.
35. Park H-J, Hong J-H, Kim H-S, et al. Predictive factors affecting cecal intubation failure in colonoscopy trainees. BMC Medical Education 2013;13:5.

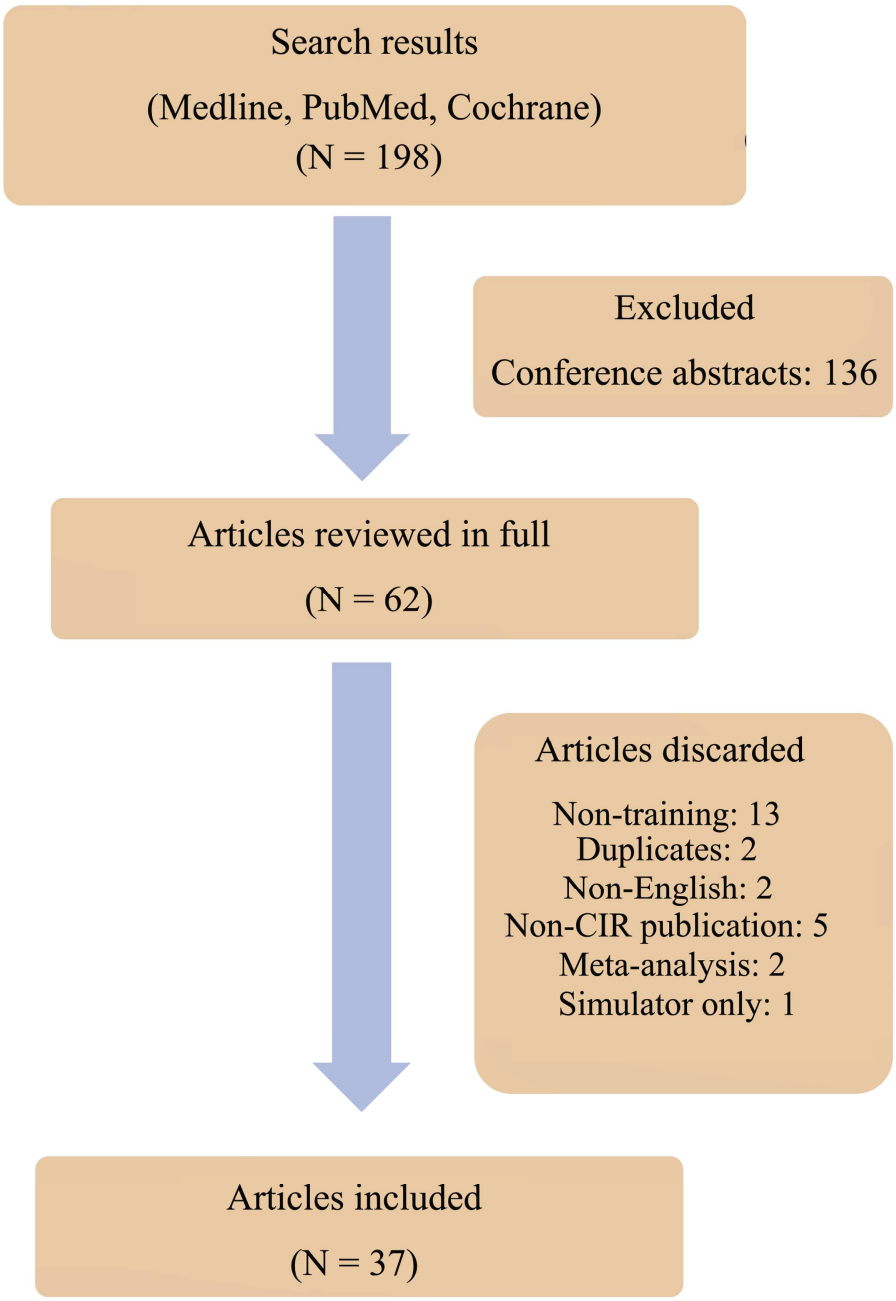
36. Luo DJ, Hui AJ, Yan KK, et al. A randomized comparison of ultrathin and standard colonoscope in cecal intubation rate and patient tolerance. Gastrointest Endosc 2012;75:484-90.
37. Manta R, Mangiavillano B, Fedeli P, et al. Hood colonoscopy in trainees: a useful adjunct to improve the performance. Dig Dis Sci 2012;57:2675-9.
38. Selvasekar CR, Holubar SD, Pendlimari R, et al. Assessment of screening colonoscopy competency in colon and rectal surgery fellows: a single institution experience. J Surg Res 2012;174:e17-23.
39. Portocarrero DJ, Che K, Olafsson S, et al. A pilot study to assess feasibility of the water method to aid colonoscope insertion in community settings in the United States. J Interv Gastroenterol 2012;2:20-22.
40. Park SM, Lee SH, Shin KY, et al. The cap-assisted technique enhances colonoscopy training: prospective randomized study of six trainees. Surg Endosc 2012;26:2939-43.
41. van Putten PG, Ter Borg F, Adang RP, et al. Nurse endoscopists perform colonoscopies according to the international standard and with high patient satisfaction. Endoscopy 2012;44:1127-32.
42. Kaltenbach T, Leung C, Wu K, et al. Use of the colonoscope training model with the colonoscope 3D imaging probe improved trainee colonoscopy performance: a pilot study. Dig Dis Sci 2011;56:1496-502.
43. Spier BJ, Durkin ET, Walker AJ, et al. Surgical resident's training in colonoscopy: numbers, competency, and perceptions. Surg Endosc 2010;24:2556-61.
44. Tee HP, Corte C, Al-Ghamdi H, et al. Prospective randomized controlled trial evaluating cap-assisted colonoscopy vs standard colonoscopy. World J Gastroenterol 2010;16:3905-10.
45. Koornstra JJ, Corporaal S, Giezen-Beintema WM, et al. Colonoscopy training for nurse endoscopists: a feasibility study. Gastrointest Endosc 2009;69:688-95.
46. Leung FW, Aharonian HS, Guth PH, et al. Involvement of trainees in routine unsedated colonoscopy: review of a pilot experience. Gastrointest Endosc 2008;67:718-22.
47. Kondo S, Yamaji Y, Watabe H, et al. A randomized controlled trial evaluating the usefulness of a transparent hood attached to the tip of the colonoscope. Am J Gastroenterol 2007;102:75-81.
48. Aslinia F, Uradomo L, Steele A, et al. Quality assessment of colonoscopic cecal intubation: an analysis of 6 years of continuous practice at a university hospital. Am J Gastroenterol 2006;101:721-31.
49. Bowles CJ, Leicester R, Romaya C, et al. A prospective study of colonoscopy practice in the UK today: are we adequately prepared for national colorectal cancer screening tomorrow? Gut 2004;53:277-83.
50. Thomas-Gibson S, Thapar C, Shah SG, et al. Colonoscopy at a combined district general hospital and specialist endoscopy unit: lessons from 505 consecutive examinations. Journal of the Royal Society of Medicine 2002;95:194-97.
51. Minozzi S, Bellisario C, Senore C. Summary documents of detailed literature searches for ESGE QIC Lower GI working group. 2017. https://www.esge.com/assets/downloads/pdfs/guidelines/s_0043_103411_1_Summary_documents_of_literature_searches.pdf.
52. Marcondes FO, Dean KM, Schoen RE, et al. The impact of exclusion criteria on a physician's adenoma detection rate. Gastrointestinal Endoscopy 2015;82:668-75.
53. Thompson C, Ismail T, Radley S, et al. Conversion of colonoscopy to flexible sigmoidoscopy: an unintended consequence of quality measurement in endoscopy. Frontline Gastroenterol 2016;7:202-06.
54. Siau K, Hawkes ND, Dunckley P. Training in Endoscopy. Curr Treat Options Gastroenterol 2018;16:345-61.
55. Rutter MD, Senore C, Bisschops R, et al. The European Society of Gastrointestinal Endoscopy Quality Improvement Initiative: developing performance measures. Endoscopy 2016;48:81-9.










Search results
(Medline, PubMed, Cochrane)
(N = 198)
Excluded
Conference abstracts: 136
Articles reviewed in full
(N = 62)
Articles discarded
Non-training: 13
Duplicates: 2
Non-English: 2
Non-CIR publication: 5
Meta-analysis: 2
Simulator only: 1
Articles included
(N = 37)

	Trainee				
	A	**B**	**C**	**D**	**E**
Procedures with caecal intubation	30	90	92	47	498
Total number of procedures	50	100	100	50	500
CIR	60%	90%	92%	94%	100%
Pooled Rate	757 / 800 = **95%**				
Mean	(0.60 + 0.90 + 0.92 + 0.94 + 1.00) / 5 = **87%**				
Median	Trainee C = **92%**				

Supplementary Table 1: Comparison of the impact of averaging methods (summary statistic) on cecal intubation rate (CIR). An example, based on invented data, highlighting the impact of outliers on the different averaging methods. Note that the pooled rate is unduly influenced by Trainee E due to the large number of procedures, whereas the mean is unduly influenced by the low CIR of Trainee A.

ACRONYMS

ASGE: American Society for Gastrointestinal Endoscopy

CIR: caecal intubation rate

DIP: Drop in performance

ERS: Endoscopy Reporting Systems

ESGE: European Society of Gastrointestinal Endoscopy

JAG: Joint Advisory Group on Gastrointestinal Endoscopy

JETS: JAG Endoscopy Training System

KPI: Key performance indicator

PCC: Provisional colonoscopy certification

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta- Analyses

UK: United Kingdom