

Trainee colonoscopy quality is influenced by the independent and unobserved performance characteristics of supervising physicians



Authors

Srihari Mahadev¹, Zhezhen Jin², Benjamin Lebwohl^{2,3}, Richard M. Rosenberg³, Reuben J. Garcia-Carrasquillo³, Ivonne Ramirez³, Daniel E. Freedberg³

Institutions

- 1 Division of Gastroenterology and Hepatology, Weill Cornell Medicine, New York, New York USA
- 2 Mailman School of Public Health, Columbia University, New York, New York, USA
- 3 Division of Digestive and Liver Disease, Department of Medicine, Columbia University, New York, New York, USA

submitted 3.5.2018

accepted after revision 30.8.2018

Bibliography

DOI <https://doi.org/10.1055/a-0770-2646> |

Endoscopy International Open 2019; 07: E74–E82

© Georg Thieme Verlag KG Stuttgart · New York

ISSN 2364-3722

Corresponding author

Srihari Mahadev, MD MS, 1283 York Ave, 9th Floor, New York, NY 10065, USA

Fax: +1-888-247-2593

smahadev@post.harvard.edu

ABSTRACT

Background Endoscopy training remains an apprenticeship, and the characteristics that facilitate transfer of high quality procedural skills from role models to trainees are

unknown. We sought to determine whether unobserved supervisor performance influences the quality of colonoscopy performed by trainees, by studying how supervisors perform alone and how trainees perform while under those same supervisors.

Methods This was a retrospective cross-sectional study conducted among ambulatory adults ≥ 50 years old who underwent colonoscopy for cancer screening or polyp surveillance from 2006 to 2015 at one academic medical center. The primary exposures were the colonoscopy withdrawal time (WT) and adenoma detection rate (ADR) of supervisors while performing colonoscopies alone. The primary outcomes were the WT and ADR of trainees performing colonoscopies under supervision.

Results Data were included from 22 attending gastroenterologist supervisors, 56 gastroenterology fellow trainees, and 2777 adults undergoing 3094 colonoscopy procedures. Among all supervised colonoscopies, mean trainee WT was 12.7 minutes (SD 4.9) and trainee ADR was 33.5%. The trainee WT was 0.42 minutes longer (standard error = 0.16, $P = 0.01$) per minute increase in supervisor WT. Similarly, trainee ADR was higher under a high ADR supervisor, and the odds ratio of high compared to low supervisor ADR category was 1.28 (95%CI 1.01 – 1.62, $P = 0.04$) after adjusting for other factors.

Conclusions The unobserved performance characteristics of supervising endoscopists may influence the quality of colonoscopy performed by trainees.

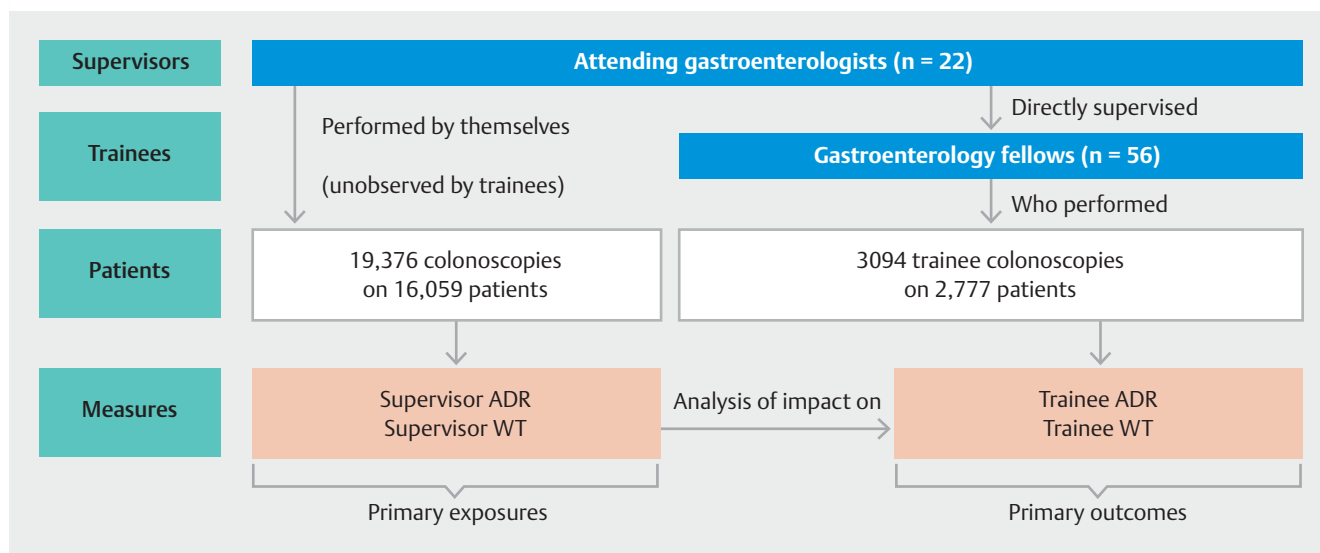
Introduction

Role modeling is fundamental to procedural training from the first shadowing experiences to the last days as a trainee [1, 2]. The paradigm of “see one, do one, teach one” pervades medical education and this paradigm begins with the observation of a more experienced practitioner, often an attending physician.

Endoscopy training, once an apprenticeship, is now recognized to merit systematic study. The reasons for this include a

broader trend towards evidence-based medicine, heightened focus on patient safety [3], increased scrutiny related to trainee independence [4], the creation in many institutions of a specialized clinical educator career trajectory, and the implementation of objective structured clinical examinations [5], video recordings [6], and simulations [7], for procedural training.

For training to improve, the role modeling paradigm must be studied [8]. While humanistic and professional qualities contribute to role model identification, it remains necessary to



► **Fig. 1** Flow chart summarizing study design, primary exposures, and primary outcomes.

identify the key factors that determine whether role models successfully inculcate high quality practices [9]. Significant variability exists in the quality of early procedural training [10]. Yet role modeling has received relatively little attention and the behaviors that determine successful role modeling remain uncertain [11].

This study aimed to understand the components of successful role modeling by determining whether unobserved supervisor characteristics influence trainee behavior. We studied supervisor and trainee performance characteristics during routine screening colonoscopies, focusing on the withdrawal time (WT) and adenoma detection rate (ADR) because they are readily quantifiable and widely recognized as quality measures for screening colonoscopy.

Patients/material and methods

Population

Data from all complete colonoscopies performed for colorectal cancer screening or polyp surveillance at a single academic medical center on patients ≥ 50 years old at average risk for colorectal cancer were included. Procedures supervised by attending physicians who infrequently performed colonoscopy (< 50 procedures performed without a trainee) or who rarely supervised colonoscopy (< 10 supervised procedures performed by a trainee) were excluded because of lack of confidence in the estimates for WT and ADR for these supervisors. This study was approved by the Institutional Review Board of Columbia University with a waiver of informed consent.

Selection of study measures

Colonoscopy WT and ADR were selected as the primary measures for the study: these metrics were both the primary exposures and the primary outcomes (► **Fig. 1**). WT is an important aspect to high quality colonoscopy [12–14] and insufficient WTs have been associated with increased risk for colorectal

cancers that arise between normal screening intervals [15, 16]. Colonoscopist ADR is generally considered to be the critical colonoscopy quality benchmark. ADR has been proposed as a reportable quality measure by the Centers for Medicare and Medicaid Services [17], and a higher provider ADR inversely correlates with patients' risk for colorectal cancer after colonoscopy [18].

Primary exposures

The primary exposures were the WT and ADR of supervisors based on procedures that they performed without trainees. To define WT in the study, data related to the time spent performing colonoscopies was retrospectively retrieved from standard endoscopy suite software (ProVation Medical, Minneapolis, Minnesota, United States). These data were generated by endoscopy suite nurses, who used electronic timestamps to document the key steps during colonoscopy. WT was then calculated as the time difference between cecal intubation and removal of the instrument at the end of the procedure. The WT calculation was restricted to the subset of studies in which no polypectomy or other maneuver was performed [15]. To define ADR in the study, data was retrieved from a pathology database and cross-referenced by patient and procedure to information retrieved from endoscopy suite software. The provider ADR was then calculated as the number of patients with adenomas divided by the number of patients screened. Supervisor WT and ADR were classified into tertiles as slow (≥ 10 minutes), medium (8 to 9.9 minutes), or fast (< 8 minutes) and as low ($< 25\%$), medium (25 to 34.9%), or high ($\geq 35\%$) for WT and ADR, respectively.

Primary outcomes

The primary outcomes were the WT and ADR of trainees based on procedures performed under the supervisors in the study. Calculation of WT and ADR is described above. Trainee WT was

► **Table 1** Characteristics of 3094 directly supervised trainee colonoscopies performed on 2777 patients.

Characteristics	n = 2777	%
Patient demographics		
Sex		
▪ Male	1172	42.2
▪ Female	1605	57.8
Age, mean (SD)		
▪ 50–59	1248	44.9
▪ 60–69	988	35.6
▪ 70–79	476	17.1
▪ ≥80	65	2.3
Procedural information		
	Number of procedures = 3094	
Indication		
▪ Screening for CRC	2357	76.2
▪ Polyp surveillance	737	23.8
Time of day		
▪ Morning procedure	1380	44.6
▪ Afternoon procedure	1714	55.4
Preparation of the bowel		
▪ Adequate	2171	70.2
▪ Inadequate	923	29.8
Withdrawal time (WT), mean min (SD) n = 313		
	12.7	(4.9)
Adenoma detection rate (ADR)		
▪ No adenomas detected	2057	66.5
▪ ≥1 adenoma detected	1037	33.5

CRC, colorectal cancer; min: minutes; SD: standard deviation.

analyzed as a continuous variable and trainee ADR as a categorical variable.

Supervision of colonoscopies by trainees

Screening colonoscopies were performed by trainees (n = 56) in their second or third/final year of fellowship training who had largely mastered basic procedural skills. All procedures were supervised in entirety by fully trained gastroenterology faculty members (n = 22), from before the moment of colonoscope insertion until after the moment of its withdrawal. At our institution, supervision is “hands-off” for trainees at this level, with recommendations or suggestions given and the supervisor taking over the procedure only for difficult maneuvers, in the case of unusually challenging procedures, or in rare cases of emergency.

Statistical approach

Automated queries were used to extract data. As co-variables, patient demographic information and procedural characteristics were retrieved electronically including the indication for the procedure (standard screening vs. polyp surveillance), preparation of the bowel (adequate vs. inadequate), and type of instrument used (standard vs. pediatric). Bowel preparation quality was derived from the endoscopist’s description within the procedure report: bowel preparation described as “poor” or “inadequate” in any segment of the colon was considered inadequate. Both the continuous measures and tertiles of the supervisor WT and ADR were also considered to be covariates. The generalized estimation equation (GEE) approach with a working independence correlation structure was used to examine the factors associated with WT using identity link and ADR using logit link, respectively. The factors which were significant at the 0.1 level in univariable GEE analysis were included in the final multivariable analysis. Sensitivity analyses were conducted to test the robustness of the observed associations. The models were rerun following exclusion of patients with inadequate bowel preparation, and on the subset of patients for whom the indication for the procedure was average-risk screening rather than surveillance due to a history of adenomas. All testing was two-sided at the alpha 0.05 level of significance. Data analysis was conducted with SAS software version 9.4 (SAS Institute Inc., Cary, North Carolina, United States).

Results

Population

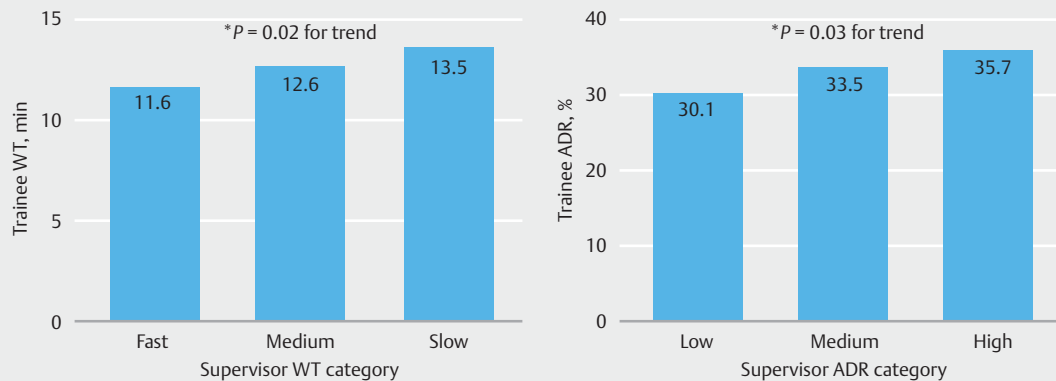
A total of 2777 patients who had 3094 colonoscopies met the criteria for inclusion in the study (► **Table 1**). The 3094 colonoscopies were performed by 56 trainees who were under the direct supervision of a group of 22 attending supervisors. Most of the procedures were for screening (76%) as opposed to surveillance (24%) and the overall adequacy of bowel preparation was 70%. Among the 3094 procedures, the trainee ADR was 33.5%. Only 313 (10.1%) of the 3094 procedures had trainee WT measurements with an average 12.7 minutes (SD 4.9).

Supervisor WT and ADR

Supervisor WT (► **Supplementary Fig. 1a**) and ADR (► **Supplementary Fig. 1b**) were classified from 19376 procedures performed without trainees. Among these procedures, mean independent supervisor WT was 8.7 minutes (SD 4.2) and ADR was 32% (95%CI 31.2–32.4).

Trainee colonoscopy characteristics

Trainee colonoscopies performed for screening (as opposed to surveillance) and those with inadequate bowel preparation were associated with slower supervisor WT category, although these associations were modest in strength. Trainee colonoscopies performed in the afternoon (as opposed to the morning) were associated with high supervisor ADR category. Again, this association was modest in strength. Trainee WT had a stepwise association with independent supervisor WT (unadjusted *P* for



► **Fig. 2** Trainee withdrawal time (WT) and adenoma detection rate (ADR), stratified by supervisor withdrawal time category and adenoma detection rate category.

trend, 0.02) and trainee ADR had a stepwise association with independent supervisor ADR (unadjusted P for trend, 0.03) (► **Fig. 2**).

Final multivariable models

In univariable GEE analysis, only supervisor WT was significantly associated with the trainee WT. The trainee WT was 0.42 minutes longer (standard error = 0.16, $P = 0.01$) per minute increase in supervisor WT. In categorized supervisor WT, trainee WT was longer under a slow withdrawing supervisor, 2.33 minutes (SE = 0.91) longer under a slow withdrawing supervisor compared to under a fast withdrawing supervisor and 1.77 minutes (SE = 0.71) longer under a medium withdrawing supervisor compared to under a fast withdrawing supervisor (► **Table 2**). The univariable GEE analysis of trainee ADR is shown in ► **Table 3**. The final model for trainee ADR included patient sex, age, and supervisor ADR or ADR category (► **Table 4**). In this model, the odds ratio of high compared to low supervisor ADR category was 1.28 (95%CI 1.01–1.62, $P = 0.04$) in adenoma detection after adjusting for other factors. Statistical significance was not met for the differences between the first two categories of supervisor ADR.

Sensitivity analyses

We performed sensitivity analyses to test the robustness of these results after excluding patients with inadequate preparation of the bowel and after excluding patients with procedures performed for surveillance rather than screening. Trends for trainee WT and ADR were similar after excluding 303 patients with inadequate preparation of the bowel (► **Supplementary Fig. 2a**) and after excluding 737 patients with procedures performed for surveillance (► **Supplementary Fig. 2b**), although the associations were no longer statistically significant. A sensitivity analysis was also performed with exclusion of outlier trainee performance, comprising the highest and lowest 5% of trainee ADR and WT. Exclusion of outliers did not influence the relationship between supervisor and trainee WT on univariable or multivariable analysis. However, the impact of supervisor

ADR category on trainee ADR no longer met statistical significance after exclusion of outliers.

Discussion

In this retrospective cross-sectional study on role modeling, the independent performance characteristics of supervisors influenced the performance characteristics of their trainees during directly supervised screening colonoscopies. A slow WT and high ADR are hallmarks of a high quality colonoscopy. We found that slow withdrawing supervisors were more likely to have trainees who withdrew slowly and high ADR supervisors were more likely to have trainees who detected adenomas. The trainees in this study were procedurally proficient, so these findings likely reflect behavioral patterns and not procedural capacity. We suspect that relatively subtle cues from supervisors tell trainees to “hurry up” or “slow down.” The independent behavior of supervisors, as measured during the procedures they performed without trainees, determined the actions of their trainees.

Although this study focused around colonoscopies done for colon cancer screening, its findings may be relevant for other scenarios. Medical trainees’ first clinical experiences are often shadowing exercises in which a supervisor models essential skills [19]. The practices learned as trainees are lifelong, with studies showing that characteristics during training tend to carry over into independent practice [20]. For those who might seek to have a private and a public face with respect to trainees, this study suggests that there is no distinction. A more positive interpretation of these findings is that encouraging good practices may create good teachers. Interventions can change the performance characteristics of attending physicians: there is high quality evidence that instituting measures that focus on increasing endoscopist ADR and WT have a direct impact upon patient outcomes including colorectal cancer mortality [21]. Our study suggests that such interventions may also improve these attending physicians as teachers or supervisors [22].

To our knowledge, no prior studies have taken a similar approach to procedural education. However, ample research af-

► **Table 2** Univariable GEE analysis for trainee withdrawal time in minutes.

Variables	Estimate (standard error)	P value
Age, per year	-0.01 (0.04)	0.65
Sex		
▪ Female	Reference	
▪ Male	-0.89 (0.57)	0.12
Bowel preparation		
▪ Adequate	Reference	
▪ Inadequate	1.47 (0.94)	0.12
Supervisor WT, min	0.42 (0.16)	0.01
Supervisor WT category		
▪ Fast	Reference	
▪ Medium	1.77 (0.71)	0.01
▪ Slow	2.33 (0.91)	0.01

GEE, generalized estimation equation; WT, withdrawal time.

► **Table 3** Univariable GEE analysis for trainee adenoma detection rate.

Variables	Odds ratio (95%CI)	P value
Age, per year	1.04 (1.03 – 1.05)	<0.01
Sex		
▪ Female	Reference	
▪ Male	1.58 (1.35 – 1.86)	<0.01
Bowel preparation		
▪ Adequate	Reference	
▪ Inadequate	0.98 (0.83 – 1.16)	0.80
Supervisor ADR (per %)	2.63 (0.91 – 7.57)	0.07
Supervisor ADR category		
▪ Low	Reference	Reference
▪ Medium	1.15 (0.93 – 1.43)	0.19
▪ High	1.29 (1.02 – 1.63)	0.03

CI, confidence interval; GEE, generalized estimation equation; ADR, adenoma detection rate.

firm the importance of role modeling. Among graduating medical students, 90% strongly identify with one or more role models and report that this interaction critically shapes clinical skills [23]. When trainees are surveyed, they are more likely to identify as excellent role models those with more time allocated towards teaching and those who emphasize the patient–doctor relationship [24, 25]. When highly regarded role models are surveyed, they also tend to stress interpersonal skills and teaching ability over personal proficiency [26]. Our findings complement these studies by suggesting that task-specific atti-

► **Table 4** Multivariable GEE analysis for trainee adenoma detection rate.

Variables	Odds ratio (95%CI)	P value
Continuous supervisor ADR		
Age, per year	1.04 (1.03 – 1.05)	<0.01
Sex		
▪ Female	Reference	
▪ Male	1.58 (1.34 – 1.85)	<0.01
Supervisor ADR (per %)	2.73 (0.94 – 7.91)	0.07
Tertiles of supervisor ADR		
Age, per year	1.040 (1.030 – 1.050)	<0.001
Sex		
▪ Female	Reference	
▪ Male	1.578 (1.343 – 1.854)	<0.001
Supervisor ADR category		
▪ Low	Reference	Reference
▪ Medium	1.121 (0.900 – 1.397)	0.307
▪ High	1.280 (1.010 – 1.622)	0.042

ADR, adenoma detection rate; CI, confidence interval; GEE, generalized estimation equation.

tudes, as well as strength or weakness as a teacher, may be an important determinant of success as a role model.

Procedural training has a mechanical or motor aspect that distinguishes it from non-procedural clinical training. The learning curve for procedures varies between individuals but learning clearly increases with experience [27–29]. Simulation has the potential to affect this learning curve [30], but role modeling is likely to remain a crucial part of procedural as well as non-procedural training. The specific quality measures assessed here are more related to attitude and approach than to innate or acquired mechanical skills, as demonstrated by studies showing that interventions can rapidly alter individuals' WT and ADR [21]. This study implies that successful role modeling of attitudes is an essential component of procedural competence, a finding in line with earlier results that personality traits strongly influence teaching in both procedural and non-procedural settings [31].

There are several strengths to this study. This is a novel approach to endoscopy education with analyses derived from data clearly linked to patient outcomes. Screening colonoscopy is a robust framework for this question because the crucial quality metrics are well established and routinely quantified. Observer bias has the potential to alter the supervisor–trainee interaction; in this study, none of the supervisors or trainees knew that their performance data was being captured and utilized to study role modeling. The study was large, covered many years, and covered hundreds of distinct supervisor – trainee pairings. Sensitivity analyses conducted on important demo-

graphic subsets demonstrated either the same associations or trends in the same direction as the primary analysis.

There are limitations to this study. It was retrospective and conducted within a single endoscopy suite at an academic medical center, which may limit its generalizability. The relationship between supervisor ADR category and trainee ADR did not meet statistical significance following exclusion of outlying trainee performers, suggesting that the results are not as robust as for the WT category; however, the trend remained unchanged. Adjustment for multiple statistical tests was not included in the analysis, raising the possibility of type 1 error. The WT data was only available for a small subset of trainee exams, which may have introduced bias. Teaching manner may have influenced trainee behavior but was not captured in the study. It is possible that supervisors were directly involved in the performance of a subset of the trainee exams. Due to the fact that the analysis involved senior fellows, however, supervisor involvement was likely limited to the insertion phase of challenging colonoscopies or to complex polypectomy maneuvers, rather than the withdrawal portion of the colonoscopy during which the colon is examined for polyps and which is most relevant for the outcome measures of WT and ADR that we chose to study.

In summary, we have shown that trainees' behavior may reflect the unobserved characteristics of supervisors while performing screening colonoscopies. Specifically, we have shown that, if the attending physician withdraws slowly while unobserved, the fellow appears to withdraw slowly. Moreover, if the supervisor finds more adenomatous polyps while unobserved, the trainee tends to find more adenomatous polyps. This suggests both that the best practitioners may make the best teachers and that private characteristics may influence the performance of trainees even when not directly observed. Role modeling is likely to remain the cornerstone of medical education in the near future. Interventions that change or modify supervisors' behaviors may improve endoscopy education, regardless of whether trainees are directly targeted by these interventions.

Acknowledgments

Dr Freedberg was funded in part by the American Gastroenterological Association (AGA) Research Scholar Award and by the National Institutes of Health (NIH K23 DK111847).

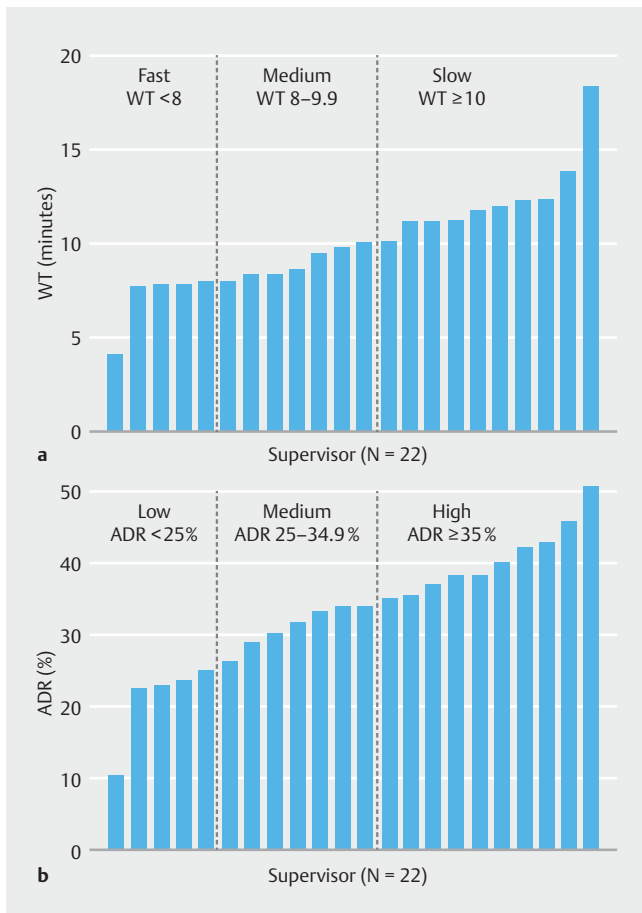
Competing interests

None

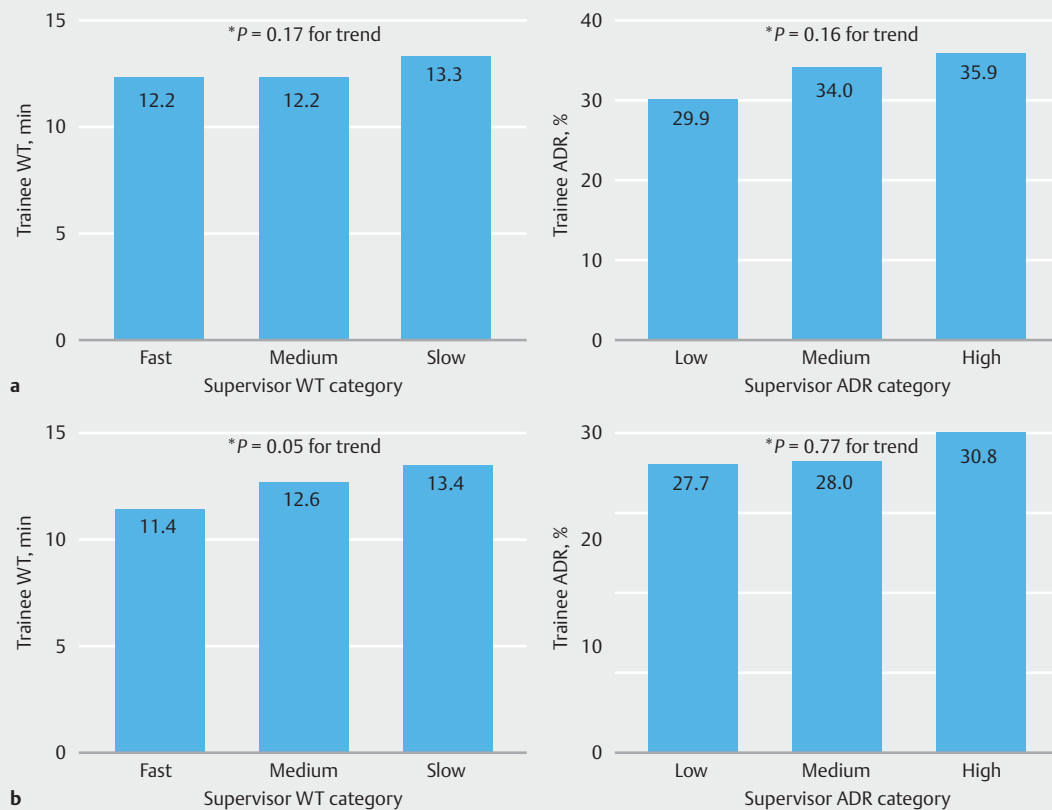
References

- [1] Sawyer T, White M, Zaveri P et al. Learn, see, practice, prove, do, maintain: an evidence-based pedagogical framework for procedural skill training in medicine. *Acad Med* 2015; 90: 1025–1033
- [2] Cruess SR, Cruess RL, Steinert Y. Role modelling – making the most of a powerful teaching strategy. *BMJ* 2008; 336: 718–721
- [3] Rodriguez-Paz JM, Kennedy M, Salas E et al. Beyond “see one, do one, teach one”: toward a different training paradigm. *Postgrad Med J* 2009; 85: 244–249
- [4] Drolet BC, Brower JP, Miller BM. Trainee involvement in patient care: A necessity and reality in teaching hospitals. *J Grad Med Educ* 2017; 9: 159–161
- [5] Harden RM, Stevenson M, Downie WW et al. Assessment of clinical competence using objective structured examination. *BMJ* 1975; 1: 447–451
- [6] Mir MA, Marshall RJ, Evans RW et al. Comparison between videotape and personal teaching as methods of communicating clinical skills to medical students. *BMJ (Clin Res Ed)* 1984; 289: 31–34
- [7] Ziv A, Wolpe PR, Small SD et al. Simulation-based medical education: an ethical imperative. *Acad Med* 2003; 78: 783–788
- [8] Benbassat J. Role modeling in medical education: the importance of a reflective imitation. *Acad Med* 2014; 89: 550–554
- [9] Ilic D, Harding J, Allan C et al. What are the attributes of a good health educator? *Int J Med Educ* 2016; 7: 206–211
- [10] Bittner JG, Coverdill JE, Imam T et al. Do increased training requirements in gastrointestinal endoscopy and advanced laparoscopy necessitate a paradigm shift? A survey of program directors in surgery *J Surg Educ* 2008; 65: 418–430
- [11] Kenny NP, Mann KV, MacLeod H. Role modeling in physicians' professional formation: reconsidering an essential but untapped educational strategy. *Acad Med* 2003; 78: 1203–1210
- [12] Rex DK, Schoenfeld PS, Cohen J et al. Quality indicators for colonoscopy. *Gastrointest Endosc* 2015; 81: 31–53
- [13] Rex DK, Johnson DA, Anderson JC et al. American College of Gastroenterology guidelines for colorectal cancer screening 2009 [corrected]. *Am J Gastroenterol* 2009; 104: 739–750
- [14] Levin B, Lieberman DA, McFarland B et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. *Gastroenterology* 2008; 134: 1570–1595
- [15] Barclay RL, Vicari JJ, Doughty AS et al. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. *NEJM* 2006; 355: 2533–2541
- [16] Shaukat A, Rector TS, Church TR et al. Longer withdrawal time is associated with a reduced incidence of interval cancer after screening colonoscopy. *Gastroenterology* 2015; 149: 952–957
- [17] Blot S, Koulenti D, Dimopoulos G et al. Prevalence, risk factors, and mortality for ventilator-associated pneumonia in middle-aged, old, and very old critically ill patients. *Crit Care Med* 2014; 42: 601–609
- [18] Corley DA, Jensen CD, Marks AR et al. Adenoma detection rate and risk of colorectal cancer and death. *NEJM* 2014; 370: 1298–1306
- [19] Kitsis EA, Goldsammer M. Physician shadowing: a review of the literature and proposal for guidelines. *Acad Med* 2013; 88: 102–110
- [20] See WA, Fisher RJ, Winfield HN et al. Laparoscopic surgical training: effectiveness and impact on urological surgical practice patterns. *J Urol* 1993; 149: 1054–1057
- [21] Kaminski MF, Wieszczy P, Rupinski M et al. Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death. *Gastroenterology* 2017; 153: 98–105
- [22] Kaminski MF, Anderson J, Valori R et al. Leadership training to improve adenoma detection rate in screening colonoscopy: a randomised trial. *Gut* 2016; 65: 616–624
- [23] Wright S, Wong A, Newill C. The impact of role models on medical students. *J Gen Intern Med* 1997; 12: 53–56
- [24] Wright SM, Kern DE, Kolodner K et al. Attributes of excellent attending-physician role models. *NEJM* 1998; 339: 1986–1993

- [25] Elzubeir MA, Rizk DE. Identifying characteristics that students, interns and residents look for in their role models. *Med Educ* 2001; 35: 272–277
- [26] Wright SM, Carrese JA. Excellence in role modelling: insight and perspectives from the pros. *CMAJ* 2002; 167: 638–643
- [27] Spier BJ, Benson M, Pfau PR et al. Colonoscopy training in gastroenterology fellowships: determining competence. *Gastrointest Endosc* 2010; 71: 319–324
- [28] Cass OW, Freeman ML, Peine CJ et al. Objective evaluation of endoscopy skills during training. *Ann Intern Med* 1993; 118: 40–44
- [29] Chak A, Cooper GS, Blades EW et al. Prospective assessment of colonoscopic intubation skills in trainees. *Gastrointest Endosc* 1996; 44: 54–57
- [30] Gallagher AG, Ritter EM, Champion H et al. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 2005; 241: 364–372
- [31] Scheepers RA, Lombarts KM, van Aken MA et al. Personality traits affect teaching performance of attending physicians: results of a multi-center observational study. *PLoS One* 2014; 9: e98107



► **Supplementary Fig. 1** **a** Withdrawal time (WT) for 22 physicians who directly supervised trainee procedures, calculated from 19 376 procedures performed without trainees. **b** Adenoma detection rate (ADR) for 22 physicians who directly supervised trainee procedures, calculated from 19 376 procedures performed without trainees.



► **Supplementary Fig. 2** Stratified analysis based on procedure indication. **a** Trainee withdrawal time (WT) and adenoma detection rate (ADR) among 2171 procedures with adequate preparation of the bowel, stratified by supervisor WT and ADR category. **b** Trainee WT and ADR among 2357 procedures performed for screening, stratified by supervisor withdrawal time and adenoma detection rate category.