

# Comparative performance and external validation of three different scores in predicting inadequate bowel preparation among Greek inpatients undergoing colonoscopy

Paraskevas Gkolfakis<sup>a,b,\*</sup>, Christina Kapizioni<sup>c\*</sup>, Georgios Tziatzios<sup>a</sup>, Antonio Facciorusso<sup>d,e</sup>, Leonardo Frazzoni<sup>f</sup>, Konstantinos Thomopoulos<sup>g</sup>, Spyros Potamianos<sup>h</sup>, Dimitrios Christodoulou<sup>i</sup>, Vasilios Papadopoulos<sup>j</sup>, Lorenzo Fuccio<sup>f</sup>, Cesare Hassan<sup>k</sup>, Konstantinos Triantafyllou<sup>a</sup>

Medical School, National and Kapodistrian University of Athens, "Attikon" University General Hospital, Athens, Greece; Erasme University Hospital, Université Libre de Bruxelles, Brussels, Belgium; Addenbrooke's Hospital, Cambridge University Hospitals NHS Trust, Cambridge, United Kingdom; University of Foggia, Foggia, Italy; University Hospital of Verona, Verona, Italy; University of Bologna, IRCSS- S. Orsola-Malpighi Hospital, Bologna, Italy; University Hospital of Patras, Patras, Greece; School of Health Sciences, University of Thessaly, Larisa, Greece; School of Health Sciences, University of Ioannina, Ioannina, Greece; Koutlimbano & Triantafylleio General Hospital, Larissa, Greece; Nuovo Regina Margherita Hospital, Rome, Italy

## Abstract

**Background** Predictive scores aim to predict bowel preparation adequacy among hospitalized patients undergoing colonoscopy. We evaluated the comparative efficacy of these scores in predicting inadequate bowel cleansing in a cohort of Greek inpatients.

**Methods** We performed a *post hoc* analysis of data generated from a cohort of inpatients undergoing colonoscopy in 4 tertiary Greek centers to validate the 3 models currently available (models A, B and C). We used the Akaike information criterion to quantify the performance of each model, while Harrell's C-index, as the area under the receiver operating characteristics curve (AUC), verified the discriminative ability to predict inadequate bowel prep. Primary endpoint was the comparison of performance among models for predicting inadequate bowel cleansing.

**Results** Overall, 261 patients—121 (46.4%) female, 100 (38.3%) bedridden, mean age 70.7±15.4 years—were included in the analysis. Model B showed the highest performance (Harrell's C-index: AUC 77.2% vs. 72.6% and 57.5%, compared to models A and C, respectively). It also achieved higher performance for the subgroup of mobilized inpatients (Harrell's C-index: AUC 72.21% vs. 64.97% and 59.66%, compared to models A and C, respectively). Model B also performed better in predicting patients with incomplete colonoscopy due to inadequate bowel preparation (Harrell's C-index: AUC 74.23% vs. 69.07% and 52.76%, compared to models A and C, respectively).

**Conclusions** Predictive model B outperforms its comparators in the prediction of inpatients with inadequate bowel preparation. This model is particularly advantageous when used to evaluate mobilized inpatients.

**Keywords** Colonoscopy, bowel preparation, predictive score

*Ann Gastroenterol* 2023; 36 (x): 1-7

Correspondence to: Georgios Tziatzios, MD, PhD, Hepatogastroenterology Unit, Second Department of Internal Medicine - Propaedeutic, Medical School, National and Kapodistrian University, Attikon University General Hospital, 1, Rimini Street, 124 62 Athens, Greece, e-mail: g\_tziatzios@yahoo.gr

Received 17 May 2022; accepted 23 September 2022; published online 15 November 2022

DOI: <https://doi.org/10.20524/aog.2023.0757>

\*These authors contributed equally to this work  
Conflict of Interest: None

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms

## Introduction

Adequate bowel preparation has been stressed as one of the most important predictors of high-quality colonoscopy. It leads to a better adenoma detection rate, which is the most important quality benchmark of colonoscopy and is inversely associated with mortality from colorectal cancer [1]. Proper surveillance intervals, improved patient safety and satisfaction, as well as mitigation of healthcare costs, represent additional advantages [2]. However, inadequate bowel preparation has been reported in 18-35% of colonoscopies, whereas current guidelines set the goal of adequate bowel preparation for ≥90% of the examinations [3]. For inpatients, the rates of inadequate bowel preparation may reach up to 50-70%, undermining the

quality of endoscopy [3]. This deviation from the recommended quality measures triggered numerous scientific attempts to identify the predictors of inadequate bowel preparation. Demographic characteristics (sex, educational level), hospitalization, comorbidities (diabetes, stroke, dementia), and medications (opioids, tricyclic antidepressants) are among the most common [4,5]. Efforts to develop a score that will easily and accurately predict inadequate bowel preparation have been undertaken, since such a score would allow the implementation of strategies to prevent it. That could have a direct impact on improving colonoscopy outcomes as well as on efficacy and cost savings for health care services. To date, 3 scores aiming to predict bowel preparation adequacy among patients undergoing colonoscopy are available, of which only one was derived exclusively from hospitalized patients. Our aim was to evaluate the comparative efficacy of the available predictive scores for bowel preparation adequacy in a cohort of Greek inpatients.

## Materials and methods

### Study design

Currently, 3 scores/models are available in the literature and thus considered eligible for the analysis (model A proposed by Dik *et al* [6], Model B proposed by Fuccio *et al* [7], and model C proposed by Gimeno-García *et al* [8]). For 2 of them (models A and C [6,8]), the cutoff value that most accurately predicts the inadequate bowel preparation is provided, while for the third one (model B [7]) an online calculator predicting the probability of inadequate bowel preparation for each patient is available. For the purposes of this study, the raw data of a

previous prospective randomized (1:1), 2 strata (mobilized vs. bedridden; 3:2) trial of consecutive inpatients from 4 tertiary centers in Greece, evaluating the role of simple and specific verbal instructions in improving the bowel preparation of inpatients (mobilized and bedridden, defined as in previous iterations [9,10]) undergoing colonoscopy were assessed to calculate the respective values for each score/model [10]. The protocol of the study was approved by the Attikon University General Hospital Ethics Committee (EBA 1677/6-7-16) and is available in the ClinicalTrials.gov registry (NCT02887014).

Patients enrolled in the original validation cohort received bowel cleansing with 4 L of polyethylene glycol solution along with a 2-day low-fiber diet, followed by a clear liquid diet the day before the colonoscopy. Bowel preparation was administered either as split doses or as same day (day before) dosing, according to the time of the colonoscopy (morning or afternoon examinations), while all examinations were performed 4-6 h after the intake of cathartics was completed [10]. Each model's compartments and cutoff values, as well as the main characteristics of their derivation/validation cohort and their efficacy in predicting the quality of bowel preparation are depicted in Table 1.

### Study outcomes

The primary outcome was to compare the performance of the available scores predicting inadequate bowel cleansing among inpatients undergoing colonoscopy. Adequate bowel preparation was defined as a total Boston Bowel Preparation Scale  $\geq 6$  with no colonic segment receiving  $< 2$  [11]. The secondary outcome was a comparison of the models' performance in predicting incomplete colonoscopies due to inadequate bowel preparation. For the primary outcome, we aimed to perform a secondary analysis of the models' performance based on the patients' stratification (mobilized versus bedridden) to check for a potential confounding effect of the patients' mobility status.

### Statistical analysis and sample size estimation

Previous studies have pointed out that substantial sample sizes are required for external validation studies [12]; therefore, as a "rule of thumb", it is suggested that at least 100 events should be available in the validation data [13], with at least 100 events and 100 non-events in the validation dataset for binary outcomes [14]. This approach is supported by simulation studies [15], assuming equal outcome prevalence in the development and validation datasets. Based on the suggestion that at least 100 events and 100 non-events for statistical tests have "reasonable power" in an external sample [16], a sample size of 261 patients would be deemed adequate [10]. Harrell's C-index, expressed as area under the receiver operating characteristics curve (AUC), was calculated to verify the accuracy of each predictive model. Comparison of the performance among the different predictive models was performed using DeLong's

<sup>a</sup>Hepatogastroenterology Unit, Second Department of Internal Medicine, Medical School, National and Kapodistrian University of Athens, "Attikon" University General Hospital, Athens, Greece (Paraskevas Gkolfakis, Georgios Tziatzios, Konstantinos Triantafyllou<sup>a</sup>); <sup>b</sup>Department of Gastroenterology, Hepatopancreatology and Digestive Oncology, Erasme University Hospital, Université Libre de Bruxelles, Brussels, Belgium (Paraskevas Gkolfakis); <sup>c</sup>Department of Gastroenterology, Addenbrooke's Hospital, Cambridge University Hospitals NHS Trust, Cambridge, United Kingdom (Christina Kapizioni); <sup>d</sup>Gastroenterology Unit, Department of Surgical and Medical Sciences, University of Foggia, Foggia, Italy (Antonio Facciorusso); <sup>e</sup>Department of Medicine, Gastroenterology and Digestive Endoscopy Unit, The Pancreas Institute, University Hospital of Verona, Verona, Italy (Antonio Facciorusso); <sup>f</sup>Department of Medical and Surgical Sciences, University of Bologna, IRCSS- S. Orsola-Malpighi Hospital, Bologna, Italy (Leonardo Frazzoni Lorenzo Fuccio); <sup>g</sup>Division of Gastroenterology, Department of Internal Medicine, University Hospital of Patras, Patras, Greece (Konstantinos Thomopoulos); <sup>h</sup>Department of Gastroenterology, Faculty of Medicine, School of Health Sciences, University of Thessaly, Larisa, Greece (Spyros Potamianos); <sup>i</sup>Division of Gastroenterology, University Hospital & Faculty of Medicine, School of Health Sciences, University of Ioannina, Ioannina, Greece (Dimitrios Christodoulou); <sup>j</sup>Department of Gastroenterology, Koutlimbanoio & Triantafylleio General Hospital, Larissa, Greece (Vasilios Papadopoulos); <sup>k</sup>Gastroenterology Unit, Nuovo Regina Margherita Hospital, Rome, Italy (Cesare Hassan)

**Table 1** Main characteristics of the assessed scores

Predictive model	Model A - Dik <i>et al</i> [6]	Model B - Fuccio <i>et al</i> [7]	Model C - Gimeno-Garcia <i>et al</i> [8]
Cohort characteristics			
No of patients enrolled	1331	1524	667
Colonoscopy indications	mixed*	mixed*	mixed *
Screening, %	27.3	0	33.2
Inpatients, %	7	100	0
Variables included in the model	ASA**** score $\geq 3$ Antidepressants (tricyclic) use Opioids Diabetes mellitus Chronic constipation History of intra-abdominal/pelvic surgery Current hospitalization History of inadequate bowel preparation	Same day regimen 2 or 4-liter PEG** solution Bedridden status Constipation Diabetes mellitus Antipsychotic use $\geq 7$ days of hospitalization	Comorbidity*** Antidepressants use Constipation History of abdominal or pelvic surgery
Cutoff	2	****	1.225
Interpretation	Score $\geq 2$ predicted inadequate bowel preparation with sensitivity 66%, specificity 79%, PPV 95% and NPV 95%	Mean predicted probability of inadequate bowel preparation: 47%	Score $\geq 1.225$ predicted inadequate bowel preparation with sensitivity 60.3%, specificity 75.4%, PPV 36.4% and NPV 89.1%

\*Including screening and diagnostic indications; \*\*polyethylene glycol-electrolyte; \*\*\*including diabetes, neurological diseases and liver cirrhosis; \*\*\*\*American Society of Anesthesiologists Physical Status Classification System; \*\*\*\*\*a cutoff is not provided, however an online calculator is available to introduce the abovementioned variables and calculate the probability of inadequate bowel preparation

PEG, polyethylene glycol; NPV, negative predictive value; PPV, positive predictive value

test. In addition, the following parameters were calculated as measures of the scores' performance:

1. Bayesian Information Criterion (BIC), which provides an estimate of the performance normalized with respect to the complexity (i.e., number of variables) of the model, where the lower the value, the better the model's predictive ability.
2. R2 coefficient of determination, which corresponds to the proportion of the variance in the dependent variable explained (i.e., predicted) by the independent variables (the predictors), both raw and adjusted for the number of variables in the model. For the R2 coefficient, the higher the value, the better the model's predicative ability.
3. Root mean squared error (RMSE), which measures the average model's error in predicting the outcome of interest, and the sigma or Residual Standard Error (RSE), that comprises a variant of the RMSE adjusted for the number of predictors in the model. For both RMSE and RSE, the lower the values, the better the model's predicative ability.
4. Performance score, a composite measure ranging from 0-100%, where higher values indicate better model performance. Calculation is based on normalizing all accuracy indices (logarithmic, quadratic/Brier and spherical score) and taking the mean value of all indices for each model.

The 3 tested models were calibrated on the study sample to ascertain the agreement between the estimated and observed number of events [17]. The main assumptions about the normality of the distribution of the residuals, collinearity of the variables used to build the model, and homoscedasticity were verified for each predictive model through visual inspection of the relevant plots. All statistical analyses were performed using the R package performance (Foundation for Statistical Computing, Vienna, Austria).

## Results

### Patients' characteristics

Overall, 261 patients—121 (46.4%) female, 100 (38.3%) bedridden, mean age  $70.7 \pm 15.4$  years—were included in the analysis. Of these, 89 (34.1%) had an inadequate bowel preparation (Boston Bowel Preparation Scale  $< 6$ ), while in 20 (7.7%) the colonoscopy was interrupted because of inadequate bowel preparation (Table 2 summarizes the patients' clinical and demographic data). All parameters—presence of comorbidities (diabetes mellitus, neurological diseases, liver cirrhosis), history of chronic constipation, history of abdominal/pelvic surgery, history of previous inadequate bowel preparation, mobility status, length of hospitalization, schema of preparation administered, use of medications (antipsychotics, antidepressants, opioids)—necessary to calculate the evaluated scores/model were available in the case report forms of the original study and were retrospectively assessed.

### Primary endpoint: comparative performance of the 3 scores/models in predicting patients with inadequate bowel preparation

Using the proposed cutoffs, models A, B and C were able to accurately detect 68 (76.4%), 87 (97.7%) and 27 (30.3%) of the 89 patients with inadequate bowel preparation. As shown in Table 3, the model proposed in model B showed both the highest performance score (100% vs. 58.49% and

0%, compared to models A and C, respectively) and Harrell's C-index values: AUC 77.2%, 95% confidence interval (CI) 75.2-78.8 vs. 72.6%, 95%CI 70.5-74.2, for model A,  $P=0.05$ ; and 77.2% vs. 57.5%, 95%CI 54.5-60.2 for model C,  $P=0.001$ ). The Akaike Information Criterion (AIC) and BIC of model B were lowest, indicating that this model was the most informative and had the best performance. Similarly, R2 (both raw and adjusted), RMSE and RSE were more favorable when model B was used (0.20, 0.19, 0.42 and 0.42, respectively) compared to the other 2 models (Fig. 1A,B). Finally, the calibration curve showed a lower discrepancy between the observed and

predicted numbers of events with model B in comparison to the other prognostic scores (Fig. 1C).

The subgroup analysis in mobilized patients confirmed the results of the main analysis (Supplementary Table 1). Model B achieved a higher performance score (100%) compared to the scores proposed by model A (11.85%) and model C (11.66%). Fig. 2 depicts the AUC for the 3 models as well as a comparison of the model indices among mobilized patients, confirming the superiority of model B in this group of patients ( $P=0.04$  between model B vs. A and  $P=0.001$  between model B vs. C). A subgroup analysis in bedridden patients was not possible, given the relatively low number of patients in this group.

**Table 2** Patients' baseline clinical and demographics characteristics (n=261)

Characteristics	Value
Female sex (%)	120 (46.2)
Age, years (mean±SD)	70.7±15.4
Mobility status, n (%)	
Mobil6dden	161 (61.7)
Bedridden	100 (38.3)
Colonoscopy indication, n (%)	
OGIB	115 (44.1)
IDA	65 (24.9)
Pain	15 (5.7)
Metastases of unknown tumor	15 (5.7)
Altered bowel habits	14 (5.4)
Other	37 (14.2)
ASA score, n (%)	
I	61 (23.3)
II	139 (53.3)
III	54 (20.7)
IV	7 (2.7)
Diabetes, n (%)	52 (19.9)
Chronic constipation, n (%)	31 (11.9)
History of abdominal surgery, n (%)	40 (15.3)
History of inadequate bowel preparation, n (%)	11 (4.2)
Use of tricyclic antidepressants, n (%)	17 (6.5)
Opioid use, n (%)	2 (0.8)
Preparation administration schedule, n (%)	
Same day	133 (51)
Split dose	128 (49)

OGIB, occult gastrointestinal bleeding; IDA, iron deficiency anemia; SOC, standard of care; ASA, American Society of Anesthesiologists physical status; SD, standard deviation

### Secondary outcome: comparative performance of the 3 scores/models in predicting patients with incomplete colonoscopy due to inadequate bowel preparation

Model B achieved the highest performance score (100%) in predicting patients with an incomplete colonoscopy due to inadequate bowel preparation compared with the scores proposed by model A (34.76%) and model C (8.99%) (Supplementary Table 2). The AUC values for the 3 models were 74.23%, 95%CI 72.5-78.4, 69.07%, 95%CI 67.4-72.1, and 52.76%, 95%CI 49.6-55.3, respectively (Fig. 3;  $P=0.02$  and  $P=0.001$  for comparisons between model B vs. A and between B vs. C, respectively). Similar to the primary endpoint, model B had the lowest AIC and BIC values, confirming that this model is the one that provides the greatest amount of information about the outcome of interest.

### Verification of model assumptions

No evidence of heteroscedasticity nor of collinearity of model parameters was observed in model B, which also was confirmed to present a normal distribution of residuals (Supplementary Fig. 1). On the other hand, normality of residuals could not be confirmed in model C (Supplementary Fig. 2), whereas collinearity and homogeneity of variance were excluded in case of model A (Supplementary Fig. 3).

### Discussion

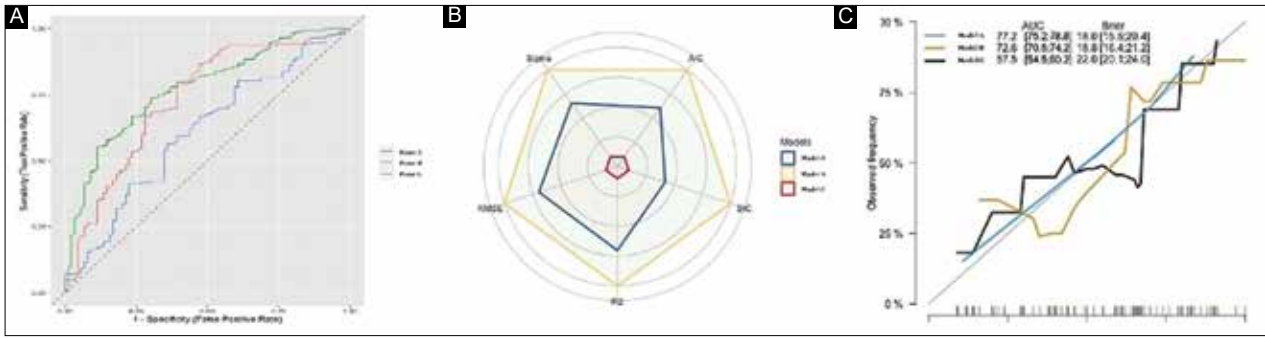
Achieving adequate bowel preparation among hospitalized patients scheduled to undergo colonoscopy can

**Table 3** Variables assessing the comparative efficacy of the 3 scores/models to predict inadequate bowel preparation in the entire cohort

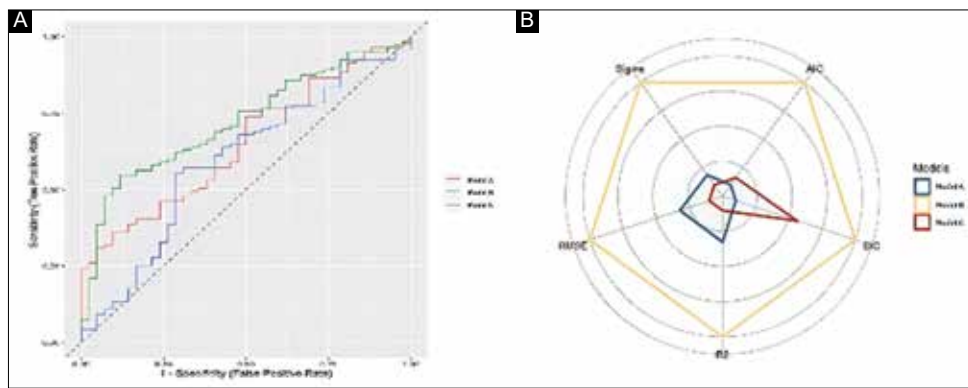
Model	AIC	BIC	R2	R2 (adjusted)	RMSE	Sigma	Performance score	AUC (95%CI)†
Model A	327.462	359.543	0.147	0.124	0.438	0.445	58.49%	72.6% (70.5-74.2)
Model B	305.864	334.380	0.209	0.190	0.422	0.427	100%	77.2% (75.2-78.8)
Model C	355.651	373.474	0.020	0.009	0.469	0.473	0%	57.5% (54.5-60.2)

†:  $P=0.05$  between Model B vs. A and  $P=0.001$  between model B vs. C

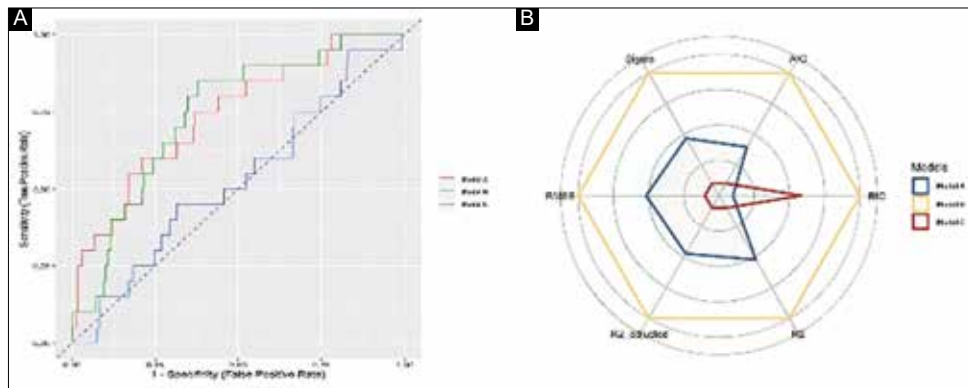
AIC, Akaike information criterion; BIC, Bayesian information criterion; R2, coefficient of determination; RMSE, root mean squared error; AUC, area under the curve; CI, confidence interval



**Figure 1** Comparative performance of the 3 scores/model for the prediction of patients with inadequate bowel preparation in the entire cohort. (A) Comparison of independent receiver operating characteristics curves for the 3 scores/model. (B) Kiviati diagram consisting of a sequence of equiangular spokes, called radii, with each spoke representing one of the variables. The data length of a spoke is proportional to the magnitude of the variable in the model. (C) Calibration curve for the 3 scores/models



**Figure 2** Comparative performance of the 3 scores/model for the prediction of patients with inadequate bowel preparation among mobilized patients (A) Comparison of independent receiver operating characteristics curves for the 3 scores/models. (B) Kiviati diagram



**Figure 3** Comparative performance of the 3 scores/model for the prediction of patients with an incomplete colonoscopy due to inadequate bowel preparation in the entire cohort (A) Comparison of independent receiver operating characteristics curves for the 3 scores/models. (B) Kiviati diagram

be a cumbersome task, with significant consequences for patients (missed diagnosis, need to repeat the examination, exposure to adverse events) as well as healthcare systems [18,19]. In an effort to address this issue, predictive models have been implemented, aiming to accurately detect cases where standard preparation practices are expected to fail and thus additional measures can be pursued. Although presently available models share several common variables, it should be stressed that they also have fundamental differences

(i.e., population evaluated) that may ultimately affect their performance. To the best of our knowledge, this is the first study to delineate an external validation of all available multivariate models for predicting inadequate bowel cleansing across Greek hospitalized patients. Our analysis showed that the predictive model proposed by Fuccio *et al* [7] (model B) significantly outperforms its predecessors (models A and C proposed by Dik *et al* [6] and Gimeno-Garcia *et al* [8], respectively). Moreover, we showed that this novel,

alternative model is particularly advantageous when used to evaluate mobilized inpatients.

Several lines of evidence pointed to inpatient status as an independent factor associated with an almost 2-fold higher risk of failed bowel preparation [5,19]. Although the reasons for this have not been studied sufficiently, hospitalized patients are generally considered prone to failing bowel preparation, as they are often of advanced age, debilitated and with severe comorbidities, factors held accountable for hampering ingestion of the desired preparation regimen volume, as well as comprehension of and compliance with instructions that may be complex. More ominous, however, is the fact that only two thirds of these patients will eventually achieve their goal, despite various interventions (education of patients and/or personnel, modification of preparation regimens) [4,10]. Aside from the inpatient status itself, there are also factors related to the setting and the individual patient (i.e., longer hospitalization time, use of antidepressants or opioids, diabetes mellitus, chronic constipation, history of intra-abdominal/pelvic surgery) that may influence the final colon cleansing result among hospitalized patients [18,19]. These are usually linked to overall impaired bowel motility, manifesting as constipation, and have been uniformly integrated in all 3 predictive models [6-8]. However, Fuccio *et al* [7], in their observational, prospective multicenter study, went a step further, also providing insights into preparation-related factors (i.e., same day preparation schedule, 2- or 4-L polyethylene glycol-electrolyte regimen) that may have the potential to increase the risk of inadequate colon cleansing. A point that deserves attention is the fact the bedridden status was highlighted as one of the strongest independent predictors of inadequate bowel preparation, a finding corroborated in a subsequent randomized controlled trial, where provision of specific verbal instructions had no beneficial impact in this particular subset of patients [10].

Our analysis demonstrated the superiority of model B compared to the 2 other comparators for identifying inpatients with inadequate preparation. This is might be a consequence of the fact that this study [7] enrolled exclusively hospitalized patients, while in 2 other previous studies the percentage of such patients was either extremely low (5.7%) or absent. This is of critical importance, given the well-established difference in bowel preparation adequacy between out- and inpatients [20,21]. Another plausible explanation is that not only were setting- and patient-related factors included as variables, but also those related to the preparation procedure itself (type/dosage/time of bowel preparation). It should be noted that the bowel cleanliness protocols applied among participating centers in each study were not only different in terms of regimen type and timing of administration (split-dose/same dose), but were also based on the discretion of the referral physician. Notably, in the study by Dik *et al* [6] patients with a previous history of inadequate bowel preparation were also included. Aside from a more comprehensive overview of each patient's probability to fail adequate colon cleansing, this also gives a pragmatic perspective, since these are perhaps the only factors than can actually be modified to

address the problem. Moreover, the study for model B was conducted across 12 different institutions where distinct preparation protocols apply, a finding that reflects regional and/or organizational policies, but also more accurately reports on real-world effectiveness.

One might repudiate the implementation of predictive models for determining the risk of inadequate colon cleansing, as they could be complex and time-consuming; however, it should be noted that all 3 models can be easily assessed, as they are straightforward and demand nothing but basic information from the patients' medical history. Indicative of the model's "operator-friendly" character is also the fact that for one of them an electronic application providing a prediction chart is available [7].

From the clinician's point of view, inadequate colon cleansing among inpatients is a commonly encountered problem. Although current European Society of Gastrointestinal Endoscopy guidelines found "*insufficient data to recommend the use of specific predictive models for inadequate bowel preparation in clinical practice*" [3], this study provides further solid evidence that adoption of a validated and easy-to-use predictive model is an efficacious intervention that could indeed assist clinicians to promptly identify patients at high risk of failing bowel preparation. This assessment can be performed accurately, regardless of the physician's level of expertise, previous dedicated training or access to the patient's medical file, while at the same time it can be replicated in other diverse settings. More importantly, this is a tool that could eventually improve the quality of bowel preparation in a setting of hospitalized patients undergoing colonoscopy, since it gives clinicians a valuable hint as to when action is necessary, while being an app for easy everyday use. Indeed, accurately predicting colonic bowel preparation adequacy before colonoscopy may facilitate the timely implementation of actions, such as systematic introduction of dose-splitting or mixed preparations, for those with many risk factors (i.e., addition of prokinetics or prolonged laxative use). Beyond any doubt, identifying a high-risk patient for inadequate colon cleansing is merely the first step in the process, while measures to deal with this situation effectively still remain to be elucidated. Nevertheless, evidence suggests that bowel cleansing in this subgroup of patients should be approached with manifold, combined strategies on a case-by-case basis, with tailored approaches that adjust the intervention to each individual patient's characteristics [4,22].

The principal strength of this study relies on its novelty, as it enhances the current bibliography by providing a validation of available models for predicting inadequate bowel cleansing among hospitalized patients. Second, our analysis included data from 2 divergent populations (mobilized/bedridden) and as such it meticulously replicates everyday clinical practice conditions. Third, an external validation study should generally take into account a slightly different case-mix to appreciate a model's portability [23]. Hence, the diversity of the population included in our study ensures the model's generalizability to a broader but relevant population and in a new independent setting; this, too could be considered a study asset.

There are also limitations of this study that merit attention. First, the low number of bedridden patients prevented an accurate analysis of this subset of patients. For the same reason, the results of the subgroup analysis regarding the mobilized inpatients should be viewed cautiously; hence, solid conclusions cannot be drawn about the performance of the aforementioned predictive models in this most difficult of all populations. Second, it should be also underlined that the included trials may have suffered from selection bias (related to physician- and patient-related preferences), an intrinsic limitation of such studies; thus, the results should be viewed in this light. Third, the sample size estimation was based on the assumption that the outcome prevalence would be equal in both the development and validation datasets. Fourth, this is a retrospective analysis; however, it highlights several areas for future research. In this regard, forthcoming prospective studies specifically designed to address this issue should be conducted, aiming to comprehensively assess the risk factors for colonoscopy preparation failure (e.g., comorbidities, medications) that could allow for identification of heterogeneity of treatment effect and define the optimal approach. Moreover, the need for validation of these scores in other patient settings is also pertinent.

To conclude, our analysis validated all available predictive models for predicting inadequate bowel cleansing in hospitalized patients, with this beneficial effect being more profound when model B was used. The same model also achieves at least equivalent performance when used to evaluate mobilized inpatients. Further trials are warranted to determine the value of such models in terms of everyday clinical practice, and to clarify strategies that could optimize bowel preparation in particular subgroups of inpatients (bedridden).

### Summary Box

#### What is already known:

- Inadequate bowel preparation before colonoscopy is frequent among inpatients
- Predictive scores aiming to predict bowel preparation adequacy among hospitalized patients undergoing colonoscopy have been introduced
- There is a lack of data comparing the efficacy of these scores for predicting inadequate bowel cleansing among Greek inpatients

#### What the new findings are:

- Model B proposed by Fuccio *et al* outperforms its comparators regarding prediction of inadequate bowel preparation in inpatients
- This predictive model is particularly advantageous for the evaluation of mobilized inpatients

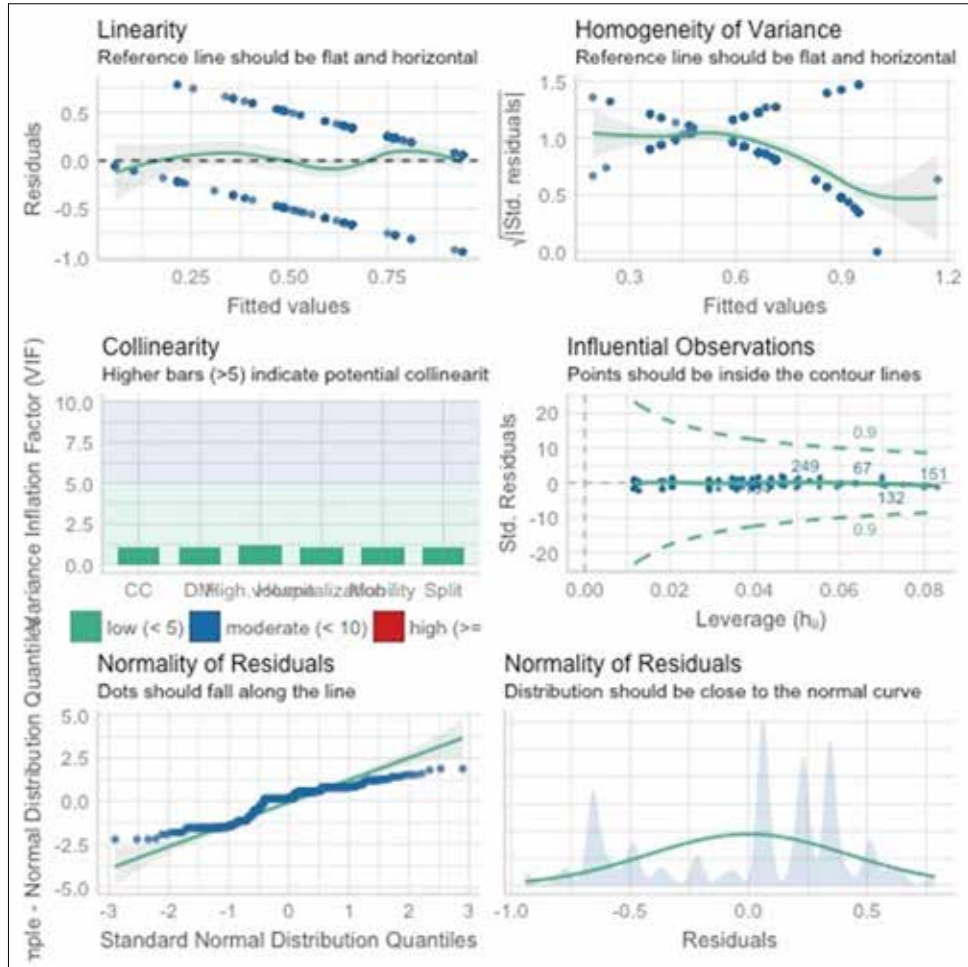
### References

1. Corley DA, Jensen CD, Marks AR, et al. Adenoma detection rate and risk of colorectal cancer and death. *N Engl J Med* 2014;**370**:1298-1306.
2. 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**:378-397.
3. Hassan C, East J, Radaelli F, et al. Bowel preparation for colonoscopy: European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2019. *Endoscopy* 2019;**51**:775-794.
4. Gkolfakis P, Tziatzios G, Papanikolaou IS, Triantafyllou K. Strategies to improve inpatients' quality of bowel preparation for colonoscopy: a systematic review and meta-analysis. *Gastroenterol Res Pract* 2019;**2019**:5147208.
5. Mahmood S, Farooqui SM, Madhoun MF. Predictors of inadequate bowel preparation for colonoscopy: a systematic review and meta-analysis. *Eur J Gastroenterol Hepatol* 2018;**30**:819-826.
6. Dik VK, Moons LM, Hüyük M, et al; Colonoscopy Quality Initiative. Predicting inadequate bowel preparation for colonoscopy in participants receiving split-dose bowel preparation: development and validation of a prediction score. *Gastrointest Endosc* 2015;**81**:665-672.
7. Fuccio L, Frazzoni L, Spada C, et al. Factors that affect adequacy of colon cleansing for colonoscopy in hospitalized patients. *Clin Gastroenterol Hepatol* 2021;**19**:339-348.e7.
8. Gimeno-García AZ, Baute JL, Hernandez G, et al. Risk factors for inadequate bowel preparation: a validated predictive score. *Endoscopy* 2017;**49**:536-543.
9. Katz S, Downs TD, Cash HR, Grotz RC. Progress in development of the index of ADL. *Gerontologist* 1970;**10**:20-30.
10. Triantafyllou K, Gkolfakis P, Skamnelos A, et al. Impact of simple, specific, verbal instructions on the quality of bowel preparation in hospitalized patients undergoing colonoscopy: a multicenter randomized controlled trial. *Endosc Int Open* 2021;**9**:E378-E387.
11. Lai EJ, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. *Gastrointest Endosc* 2009;**69**:620-625.
12. Peek N, Arts DG, Bosman RJ, van der Voort PH, de Keizer NF. External validation of prognostic models for critically ill patients required substantial sample sizes. *J Clin Epidemiol* 2007;**60**:491-501.
13. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**:361-387.
14. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Habbema JD. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *J Clin Epidemiol* 2005;**58**:475-483.
15. Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;**35**:214-226.
16. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating* (Steyerberg EW, ed). Cham, Springer International Publishing, 2019.
17. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;**162**:W1-W73.
18. Garber A, Sarvepalli S, Burke CA, et al. Modifiable factors associated with quality of bowel preparation among hospitalized patients undergoing colonoscopy. *J Hosp Med* 2019;**14**:278-283.
19. Yadlapati R, Johnston ER, Gregory DL, Ciolino JD, Cooper A,

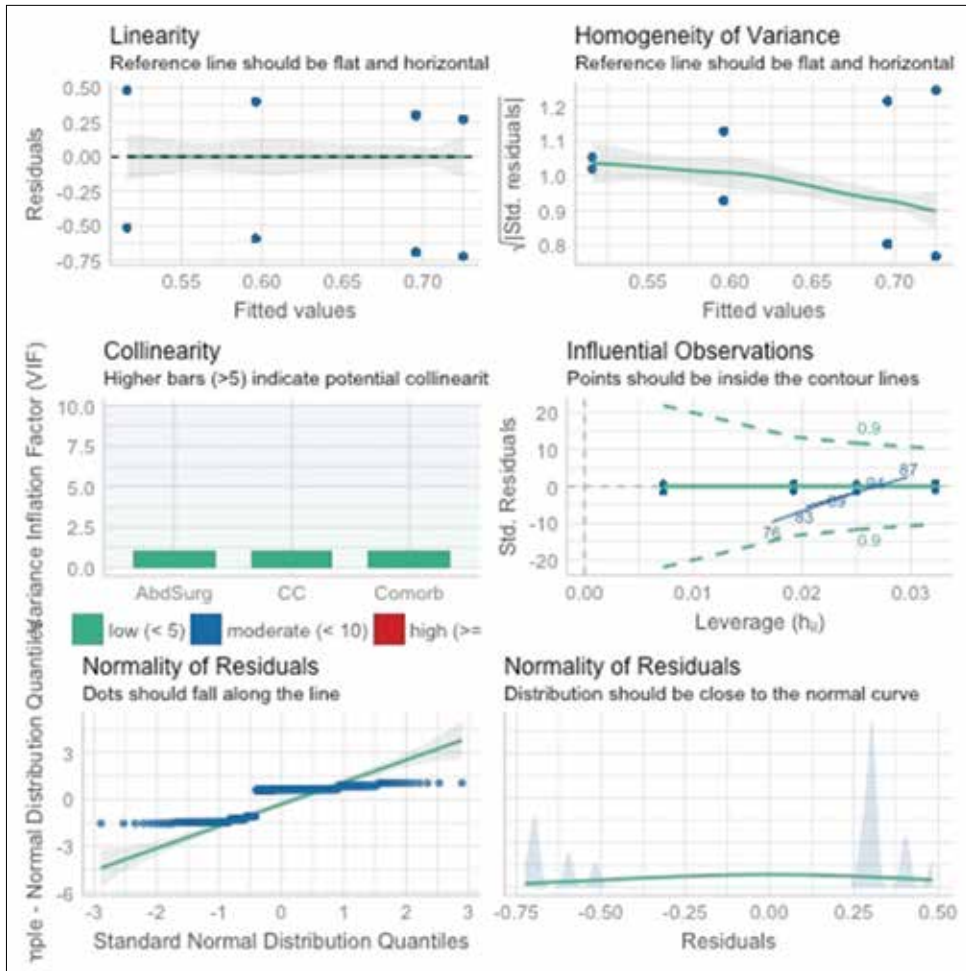
- Keswani RN. Predictors of inadequate inpatient colonoscopy preparation and its association with hospital length of stay and costs. *Dig Dis Sci* 2015;**60**:3482-3490.
20. Froehlich F, Wietlisbach V, Gonvers JJ, Burnand B, Vader JP. Impact of colonic cleansing on quality and diagnostic yield of colonoscopy: the European Panel of Appropriateness of Gastrointestinal Endoscopy European multicenter study. *Gastrointest Endosc* 2005;**61**:378-384.
21. Ness RM, Manam R, Hoen H, Chalasani N. Predictors of inadequate bowel preparation for colonoscopy. *Am J Gastroenterol* 2001;**96**:1797-1802.
22. Argyropoulos SK, Mahmood SK, Campbell EJ, Richter JM. Improving the quality of inpatient bowel preparation for colonoscopies. *Dig Dis Sci* 2018;**63**:338-344.
23. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. *J Clin Epidemiol* 2015;**68**:25-34.



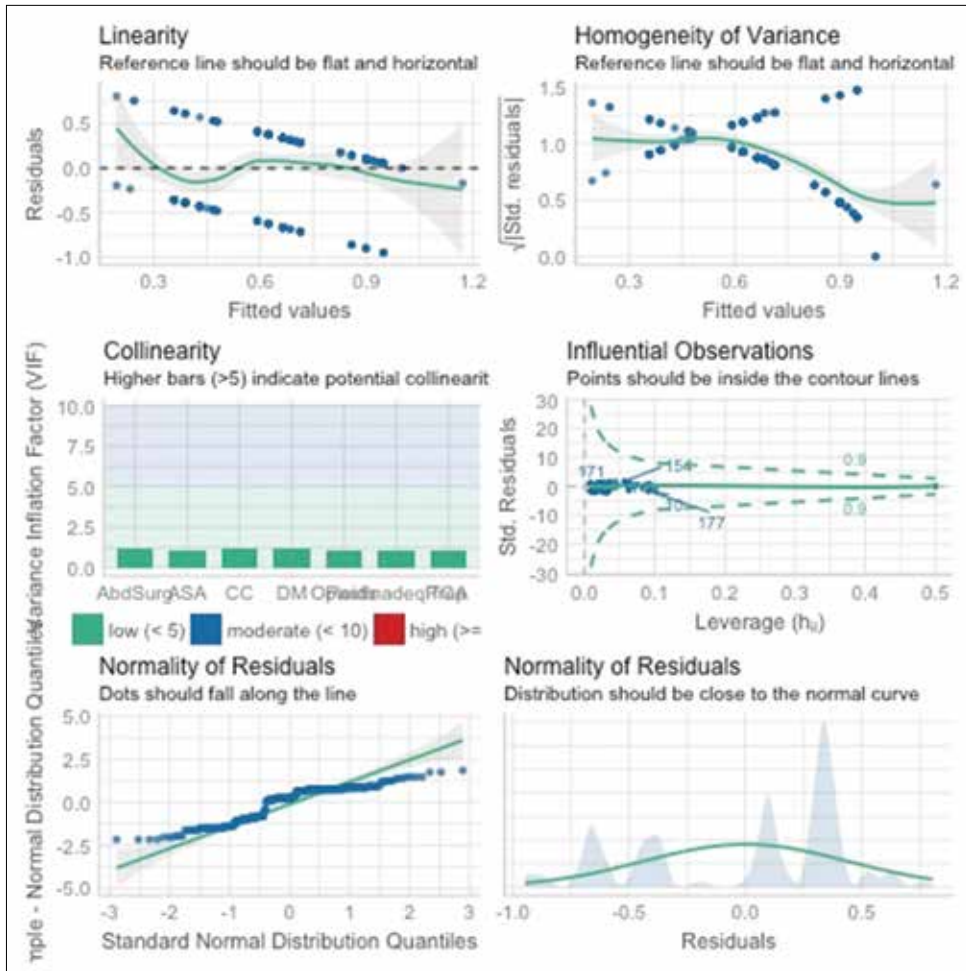
Supplementary material



Supplementary Figure 1 Verification of assumptions for Model B as proposed by Fuccio *et al* [7]. All assumptions are fulfilled



Supplementary Figure 2 Verification of assumptions for Model C as proposed by Gimeno-Garcia *et al* [8]. All assumptions are fulfilled



Supplementary Figure 3 Verification of assumptions for Model A as proposed by Dik *et al* [6]. All assumptions are fulfilled