The European Society of Gastrointestinal Endoscopy Quality Improvement Initiative: developing performance measures

Authors Matthew D. Rutter[1,2], Carlo Senore[3], Raf Bisschops[4], Dirk Domagk[5], Roland Valori[6], Michal F. Kaminski[7,8], Cristiano Spada[9], Michael Bretthauer[8,10,11], Cathy Bennett[12], Cristina Bellisario[3], Silvia Minozzi[3], Cesare Hassan[13], Colin Rees[1], Mário Dinis-Ribeiro[14], Tomas Hucl[15], Thierry Ponchon[16], Lars Aabakken[10], Paul Fockens[17]

Institutions Institutions are listed at end of article.

Bibliography
DOI http://dx.doi.org/10.1055/s-0035-1569580
Published online: 11.12.2015
Endoscopy 2016; 48: 81–89

ISSN 0013-726X

Corresponding author
Matthew Rutter, MB BS, MD
European Society of Gastrointestinal Endoscopy (ESGE)
c/o Hamilton Services GmbH
Landwehr Str. 9
80336 Munich, Germany
Fax: +49-89-907793620
secretariat@esge.com

The European Society of Gastrointestinal Endoscopy (ESGE) and United European Gastroenterology (UEG) have a vision to create a thriving community of endoscopy services across Europe, collaborating with each other to provide high quality, safe, accurate, patient-centered and accessible endoscopic care. Whilst the boundaries of what can be achieved by advanced endoscopy are continually expanding, we believe that one of the most fundamental steps to achieving our goal is to raise the quality of everyday endoscopy. The development of robust, consensus- and evidence-based key performance measures is the first step in this vision.

ESGE and UEG have identified quality of endoscopy as a major priority. This paper explains the rationale behind the ESGE Quality Improvement Initiative and describes the processes that were followed. We recommend that all units develop mechanisms for audit and feedback of endoscopist and service performance using the ESGE performance measures that will be published in future issues of this journal over the next year. We urge all endoscopists and endoscopy services to prioritize quality and to ensure that these performance measures are implemented and monitored at a local level, so that we can provide the highest possible care for our patients.

Abbreviations

ADR	adenoma resection rate
AGREE	Appraisal of Guidelines for Research and Evaluation
AMSTAR	Assessing the Methodological Quality of Systematic Reviews
ASGE	American Society for Gastrointestinal Endoscopy
CARE	Complete Adenoma Resection [study]
CIR	cecal intubation rate
CRC	colorectal cancer
EOI	expression of interest
ERCP	endoscopic retrograde cholangiopancreatography
ESGE	European Society of Gastrointestinal Endoscopy
GI	gastrointestinal
GRADE	Grading of Recommendations Assessment, Development and Evaluation
ISFU	Importance, Scientific acceptability, Feasibility, and Usability
NQMC	National Quality Measures Clearinghouse
PCCRC	post-colonoscopy colorectal cancer
PICOS	population/patient, intervention, comparison, outcome, study design
QUADAS	Quality Assessment Tool for Diagnostic Accuracy Studies
QIC	Quality Improvement Committee
SIGN	Scottish Intercollegiate Guidelines Network
UEG	United European Gastroenterology

The importance of quality

Tens of millions of people undergo endoscopic procedures every year in Europe. Endoscopy is the pivotal investigation in the diagnosis of gastrointestinal pathology and a powerful tool in its management. High quality endoscopy delivers better health outcomes and a better patient experience [1]. yet there is clinically significant variation in the quality of endoscopy currently delivered in endoscopy units [2–6].

An example of this is post-colonoscopy colorectal cancer (PCCRC). It is known that the majority of PCCRCs arise from missed lesions (premalignant polyps or cancers]. or incomplete polypectomy [7,8]. Back-to-back colonoscopy studies show that 22% of all adenomas are missed [9–14], and that there is a three- to sixfold variation in adenoma detection rates between endoscopists [15,16].

Even when polyps are found, removal may be incomplete: the Complete Adenoma REsection (CARE) study concluded that 10% of nonpedunculated polyps of 5–20 mm and 23% of nonpedunculated polyps of 15–20 mm were incompletely resected [17]. Furthermore, low cecal intubation rates and poor bowel preparation regimens may explain the relative failure of colonoscopy to protect against proximal colorectal cancer that was found in many studies [18–25]. This results in clinically important differences in quality of care and patient outcomes: a recent study in the UK demonstrated a more than fourfold variation in PCCRC rates between hospitals [26].

In the upper GI tract, gastric cancers and precursor lesions are frequently missed: in one series, 7.2% of patients with gastric cancer did not have the lesion detected at endoscopy performed in the preceding 1 year. Of these cases, almost three quarters were felt to be due to endoscopist error [27]. Equally, in ERCP, which is one of the most complex and highest risk procedures performed regularly in endoscopy practice, there is evidence of wide variation in both completion and complication rates [28–35].

Performance measures

Providers and users of services can only know whether their service is delivering good quality care if it is measured. *Performance measures* are measurements that are used to assess the performance of a service or aspect of a service; other terms used for these include *quality measures*, *quality indicators*, *key performance indicators*, or *clinical quality measures*. Evidence-based performance measures provide endoscopists and endoscopy units, both often working in relative isolation, with a framework and benchmark against which they can assess their service.

Knowledge of the significant variation in quality between endoscopists does not improve quality per se, but setting minimum and target standards within these measures incentivizes improvement: when clinicians and services see their own performance data, they act to improve them. Open publication of performance measures also permit users of the service to assess quality for themselves, thus making better informed choices and further incentivizing improvements in healthcare. However, although open publication has potential benefits, it can cause unintended damage if handled poorly, for example if data are open to misinterpretation or inappropriate comparison. Thus it is important to consider both the benefits and risks of open publication for each case.

The provision of high quality endoscopic care is complex, involving myriad people, processes, and equipment. Healthcare professionals work hard to deliver this service, yet failure of any aspect may result in suboptimal care and poor health outcomes. Performance measures help a service to identify, appraise, and monitor the key steps in the process and the key outcomes, showing where systems are suboptimal and whether the service is providing high quality patient-centered healthcare.

Carefully constructed performance measures should allow providers to identify and address specific deficits in their service, resulting in better patient outcomes. Good performance measures should therefore correlate with an important health outcome. These measures should be evidence-based, clear, objective, reproducible, and realistic. They should also be practical to measure and meaningful for their target audience (for example endoscopists, patients, or healthcare providers). In an ideal construct, there should be a small number of carefully selected performance measures assessing all important aspects of the service (domains). Each measure assesses performance from a specific angle. Together they provide a holistic snapshot of the quality of the service. Some performance measures may relate to broad procedures (for example, cecal intubation rate), whereas others may relate to specific steps in a specific procedure (for example the optimal biopsy strategy for surveillance of Barrett's esophagus).

Performance measures can be used to measure the quality of organizational structure, healthcare processes, or clinical outcomes. They can be applied in the pre-, intra- or post-procedural time periods.

- **Structural measures** reflect the conditions in which providers care for patients, in other words they reflect aspects of healthcare infrastructure. These measures can provide information about procedural volumes performed by a provider, staffing levels or, for example, whether a provider has adopted an electronic endoscopy reporting system.
- **Process measures** show whether actions proven to benefit patients are being completed. An example would be the percentage of patients requiring pre-procedure antibiotics who receive the correct antibiotic at the correct time.
- **Outcomes measures** analyze the actual results of care. These are generally the most important measures. An example would be the percentage of patients readmitted to hospital for a complication within 30 days of the endoscopic procedure.

Performance measures describe what to measure. However, it is usually desirable to take this further, identifying a minimum standard and a target standard within the measure. For example, it might be decided that cecal intubation rate is an important performance measure of colonoscopy; within this, a minimum standard might be set at 90% or 95%, with a target standard of 97%. Whereas performance measures will remain relatively static over time, the standards within such measures will be more dynamic, changing over time as techniques and technology improve. Moreover, the standards may vary according to procedure: for example, the minimum standard for adenoma detection rate will be higher for diagnostic colonoscopy performed because of fecal occult blood findings compared with colonoscopy prompted by symptoms. Occasionally no clear minimum standard currently exists for a performance measure (for example, patient comfort), yet its assessment may still be considered important. These are sometimes described as "auditable outcomes," and it is hoped that in time, further research will help determine appropriate standards. Owing to small sample size, rates for rare events, such as missed cancers, may be best examined at endoscopy unit level rather than endoscopist level, whilst a qualitative review of each case is also performed (root cause analysis).

The terminology used in measuring quality can be confusing. A summary of terminology is presented in **Table 1**.

Table 1 Terminology used in measuring quality.

Term	Description/definition	Example
Domain	An area of clinical practice	Completeness of procedure, identification of pathology, management of pathology, complications, patient satisfaction
Performance measure	A measure that helps assess performance within a domain. Other terms used for this include *quality measure*, *quality indicator*, *key performance indicator*, or *clinical quality measure*. Can look at structure, process, or outcome.	Cecal intubation rate (CIR)
Minimum standard	A minimum defined level of performance within a performance measure	Minimum CIR standard is ≥ 90 %
Target standard	A desirable/aspirational level of performance within a performance measure	Target CIR standard is ≥ 95 %

The ESGE Quality Improvement Initiative

The ESGE Quality Improvement Committee (QIC) was instigated in 2013. Its aims are:

- To improve the global quality of endoscopy and the delivery of patient-centered endoscopy services
- To promote a unifying theme of quality of endoscopy within ESGE activities, achieved by collaborating with other ESGE committees and working groups and underpinned by a clear quality improvement framework
- To assist all endoscopy units and endoscopists in achieving these standards.

QIC committee membership comprises the QIC chairperson (M.R.), ESGE president and president-elect, chairs of the other three ESGE committees (guidelines, education and research) and chairs of QIC working groups.

A QIC strategy was developed to aid fulfilment of ESGE QIC aims. Quality improvement is a dynamic process and as such the strategy details will evolve over time, although the broad quality remit will not. An initial key objective was to help improve the quality of gastrointestinal endoscopy by producing a framework of performance measures for endoscopy, including quality of independent endoscopists and quality of endoscopy services (covering all aspects of the service including equipment, decontamination, waiting times, and patient experience), by developing robust, evidence-based performance measures. The aim of this was to set a minimum standard for individual endoscopists and for the endoscopy service, and to permit endoscopy units to measure their services against this patient-centered framework.

It was determined that such performance measures should be constructed using a rigorous evidence-based consensus process, incorporating a wide variety of stakeholders, including patients, from as wide a geographical area as possible. The aim was to delineate the core domains of a quality endoscopy service, to identify performance measures within each domain, and precisely to define and describe a small number of key performance measures covering each domain.

As the project fulfilled a key aim of the UEG Strategic Plan 2015 – 2018, ESGE approached UEG regarding potential collaboration and UEG agreed to this collaboration. Both ESGE and UEG co-funded the project and provided additional project governance.

The QIC committee created four working groups related to different areas of the gastrointestinal (GI) tract: upper GI, lower GI, pancreatobiliary, and small-bowel. A fifth "Endoscopy Service" working group was also created. An open call for expressions of interest (EOI) in participation was launched by ESGE, by emailing all individual members and all ESGE-affiliated endoscopy societies and by placing an article in the ESGE newsletter. A total of 90 EOIs were received from over 30 nations. The QIC committee nominated, approached, and appointed working group chairs and a meeting with these chairs was held to discuss the project in detail. Utilizing the list of EOIs, each working group chair established their working group membership, aiming to ensure as wide a geographical spread as possible, with between 10 and 20 members per GI tract group. Because of the nature of the Endoscopy Service group with regards to varying practice between nations, membership of this working group was deliberately larger and each ESGE-affiliated national endoscopy society was asked to nominate an individual to participate in the group, which comprised 34 members. No individual was permitted to be in more than one group. The American Society for Gastrointestinal Endoscopy (ASGE) was approached regarding collaborative involvement and agreed to provide input specifically into the small-bowel working group, along with overall comment or endorsement of the project output as appropriate.

The QIC committee contracted an expert team of methodologists to provide methodological support and to conduct the detailed literature searches (Literature Group). The Literature Group leader (C.S.) was co-opted onto the QIC committee for the duration of the project. To facilitate the program, a bespoke web-based platform was commissioned (ECD Solutions, USA). Within this platform, modules were created corresponding to the steps in the development process. All working group members had access to these modules, permitting both open and anonymized discussion around each aspect of the performance measure development. An expert in guideline methodology with significant prior experience of working with similar web-based platforms (C. Bennett) was commissioned to facilitate the integration of the information technology component.

Performance measures project process

A multistep process was developed by the QIC committee (**Table 2**). The Appraisal of Guidelines for Research and Evaluation II (AGREE II) tool was used to structure the guideline development process [36], incorporating best practice from both the Scottish Intercollegiate Guidelines Network (SIGN) development processes and the National Quality Measures Clearinghouse (NQMC) of the United States of America. To ensure working group members had an understanding of guideline development methodology, all completed the SIGN online critical appraisal course (http://www.sign.ac.uk/methodology/tutorials.html; with permission).

A preliminary meeting for all working group members was held at the UEG Week conference in Vienna, October 2014. The project was explained in detail and each working group proposed potential domains for endoscopy. After open discussion, a draft single set of domains, unified across all the four GI tract areas, was constructed and voted on using a modified Delphi consensus pro-

Table 2 Performance measures project: process steps.

Establishment of QIC and project working groups
Declaration of conflicts of interest – all working group members
Complete SIGN online critical appraisal course – all working group members
Define the domains across all four GI fields (upper GI, small-bowel, pancreatobiliary, lower GI) and separately for Endoscopy Service (agreed by modified Delphi consensus process across all working groups)
Create PICOs, listing all key outcomes
Conduct literature search and construct evidence table
Create long-list of performance measures for each domain within each working group
Use ISFU checklist (**Table 5**) for each potential performance measure. Discard inferior performance measures, and where no performance measure exists within a domain, construct appropriate performance measure by modified Delphi consensus process
Determine final performance measures – modified Delphi consensus process
Develop descriptive framework for each performance measure (**Table 6**). Review, tabulate and GRADE evidence for minimum/target standards within each performance measure
Review and harmonization of performance measures across all five working groups
Highlight areas for future research based on gaps in evidence identified during this process
Identify training/education needs
Review by ESGE, UEG, national societies, and patient groups for comment and consensus
Final amendments – modified Delphi process including ESGE QIC committee

QIC, Quality Improvement Committee; SIGN, Scottish Intercollegiate Guidelines Network; GI, gastrointestinal; PICOS, population/patient, intervention, comparison, outcome, study design; ISFU, Importance, Scientific acceptability, Feasibility, and Usability; GRADE, Grading of Recommendations Assessment, Development and Evaluation; ESGE, European Society of Gastrointestinal Endoscopy : UEG, United European Gastroenterology.

Table 3 Modified Delphi consensus process.

Consensus voting was conducted through the website. Consensus was reached using a modified Delphi technique. Each working group member anonymously scored their level of agreement with draft measures using a 1 to 5 scale: 1 = Strongly agree, 2 = Agree, 3 = Neither agree nor disagree, 4 = Disagree, 5 = Strongly disagree.
Space was provided to include comments and additional references that were felt to require consideration. Commenting was mandatory for undecided or disagree votes.
At least 80 % agreement (scores of 1 or 2) was required for consensus to be reached. Where consensus was not reached, measures were reviewed in light of comments made and any additional evidence identified, and were adjusted if required. Further voting rounds then took place for these measures.
If 80 % agreement was not reached after a maximum of three rounds of voting, consensus was considered reached if > 50 % of participants voted in favor and < 20 % voted against the measure, in accordance with the GRADE process [37]. Failure to meet this criterion resulted in the measure being discarded.

cess, as described in **Table 3** [38]. If consensus was not reached initially, further discussion and voting was performed to re-evaluate and modify proposed domains until consensus was reached. The agreed domains for the GI tract working groups included completeness of procedure, identification of pathology, management of pathology, complications, procedure numbers, and patient experience.

Each working group developed an exhaustive list of potential areas for literature review, using the PICOS (Population/Patient, Intervention, Comparison, Outcome, Study design) process [39–41]. The questions were focused on the assessment of the relationship between specific indicators and procedure outcomes (e.g. completion rate) or patient outcomes (e.g. interval cancer rate, change in clinical management). PICOS were reviewed by the Literature Group and revisions made until a final precisely defined list was reached. The PICOS components of each prioritized question were used by the Literature Group to define specific keywords for the comprehensive bibliographic searches. If more than one comparison was deemed to be relevant, the results of each comparison were reported.

Searches were performed on the Cochrane Central Register of Controlled Trials (CENTRAL), Medline and Embase, from 1 January 2000 to 28 February 2015, using MESH terms and free-text words, without language restriction. In the first instance systematic reviews were searched. If updated systematic reviews addressing the PICOS questions were retrieved, the search for primary studies was limited to those studies published after the last search date of the most recently published systematic review. If no systematic reviews were found, a search of primary studies since 2000 was performed. In order to avoid repetition or double counting of primary studies, where a literature search retrieved many systematic reviews addressing the same PICOS question, only the best systematic review, based on the evaluation of their methodological quality, update of the bibliographic search, level of overlapping, and quality of evidence of included primary studies, was considered for data extraction.

A hierarchy of the study designs to be considered for each type of question (e.g. on effectiveness, diagnostic accuracy, acceptability, and compliance) was produced by the epidemiologists of the Literature Group. For effectiveness questions, randomized controlled trials were considered as the best source of evidence and were searched in the first instance. For diagnostic accuracy questions, cross-sectional studies with verification by reference standard were considered as the best source of evidence.

The risk of bias of included studies was assessed using the following validated checklists:

- systematic review: AMSTAR (Assessing the Methodological Quality of Systematic Reviews) checklist [42]
- randomized controlled trials: The Cochrane Collaboration's tool for assessing risk of bias in randomized trials [43]
- cohort studies, case-control studies and cross-sectional surveys: Newcastle-Ottawa Scale [44]
- diagnostic accuracy studies: QUADAS 2 (Quality Assessment Tool for Diagnostic Accuracy Studies 2) checklist [45]
- interrupted time series analysis: criteria suggested by the Cochrane Effective Practice and Organisation of Care Review Group [46].

The draft results of the bibliographic search and of the selection process produced by the Literature Group were reviewed by the clinical experts of the working groups, to determine whether the inclusion of additional evidence or the exclusion of nonrelevant papers was required. Once necessary revisions were made, for each question or group of questions pertaining to the same topic, the Literature Group provided an evidence table with the main characteristics of each included study (study design, objective of the study, comparisons, participant characteristics, outcome measures, results, risk of bias). They also provided a summary document with a description of the search strategy used for each database, the overall number of titles retrieved, and the

number of potentially relevant studies acquired in full text; the number of studies finally included was given, as well as a synthesis of their characteristics and risk of bias, and of their results, overall conclusions, and quality of evidence.

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool was used to evaluate both the quality of evidence and the strength of recommendations made (**Table 4**) [48, 49]. The GRADE system specifically separates the quality of evidence from the strength of a recommendation: whilst the strength of recommendation may often reflect the evidence base, the GRADE system allows for occasions where this is not the case, for example where there appears to be good reason to make a recommendation in spite of an absence of high quality scientific evidence such as a large randomized controlled trial.

Once the literature review was completed, initial draft evidence statements with comprehensive supporting documentation were uploaded onto a customized web platform, for all working group members to review and comment in a modified Delphi process (see **Table 3**), to allow modification and to identify additional references. Where necessary, further literature reviews were undertaken and further revisions made in subsequent voting rounds.

From the final evidence construct, the working group chairs identified draft performance measures, aiming for a small number of key measures per domain. Where no measure had been identified within a domain, the working group was permitted to construct one by consensus if deemed clinically appropriate. Once the key performance measures had been identified, each measure was evaluated using the ISFU (Importance, Scientific acceptability, Feasibility, and Usability) framework described by the National Quality Measures Clearinghouse (**Table 5**) [50]. Measures which did not meet the criteria were discarded. The modified Delphi process was then used to reach consensus on these performance measures.

A detailed descriptive framework was then constructed for each measure meeting the ISFU criteria, as described in **Table 6** [51]. Quality standards (minimum and target) were identified within each performance measure. Additional literature searches were performed where necessary. Where no evidence-based standard was identified, the working group was permitted either to agree on a suitable standard by consensus, or to state "no current standard defined."

Along with the final list of precisely defined key performance measures, the working groups compiled a longer list of other performance measures that had been identified during the development process, a list of areas with weak evidence base for priority research, and a list of training/educational needs. The final draft was then reviewed by the ESGE QIC Committee and the ESGE Governing Board. Finally, review and approval was obtained from ESGE-affiliated national societies, UEG, ASGE, and patient groups.

Table 4 An overview of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system [47].

GRADE: Strength of evidence
High quality: Further research is very unlikely to change our confidence in the estimate of effect
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low quality: Any estimate of effect is very uncertain
GRADE: Strength of recommendation
Recommendations can be categorized as either **Strong** or **Weak**. Recommendations involve a trade-off between benefits and harms. Those making a recommendation should consider four main factors: – The trade-offs, taking into account the estimated size of the effect for the main outcomes, the confidence limits around those estimates, and the relative value placed on each outcome – The quality of the evidence – Translation of the evidence into practice in a specific setting, taking into consideration important factors that could be expected to modify the size of the expected effects, such as proximity to a hospital or availability of necessary expertise – Uncertainty about baseline risk for the population of interest. If there is uncertainty about translating the evidence into practice in a specific setting, or uncertainty about baseline risk, this may lower our confidence in a recommendation.

The ESGE quality improvement vision

ESGE and UEG have a vision to create a thriving community of endoscopy services across Europe, collaborating with each other to provide high quality, safe, accurate, patient-centered, and accessible endoscopic care. Whilst the boundaries of what can be achieved in advanced endoscopy are continually expanding, we believe that one of the most fundamental steps to achieving our goal is to raise the quality of everyday endoscopy. The development of robust, consensus- and evidence-based key performance measures is the first step in this vision.

Implementing performance measures, along with additional measures such as structured training programs, can result in significant improvement in endoscopy quality. In the UK for example, a decade of quality improvement initiatives resulted in cecal intubation rate improving from 76.9 % to 92.3 % [52].

Having a performance measure does not result in improved health outcomes per se: in order to improve quality, it is essential to measure local performance regularly against this benchmark. Services and individuals are unlikely to improve unless they are aware of their performance and how it compares with benchmark performance measures. Measuring allows the identification of potential underperformance, which provides an opportunity for discussion and support for the endoscopist. In addition, the simple act of monitoring a service will improve performance (the "Hawthorne effect"): it is powerful, essentially free, and results in improved quality of patient care.

The standardization of performance measure definitions and measurement methodology is crucial to permit comparative assessment. Quality improvement requires political will. At a local level, it requires support from hospital management. Whilst not essential, the best examples of quality improvement in endoscopy have also had commitment from, indeed have often been led by, regional or national authorities and we call upon such organizations to share responsibility for and to facilitate this program. The implementation of appropriate information technology infrastructure, based around electronic endoscopy reporting systems, is an important step in allowing timely data collection and automated, standardized performance measure reporting.

A strong case can be made for setting a minimum number of procedures per endoscopist per year. Firstly, a large sample size in-

Table 5 Importance, Scientific acceptability, Feasibility, and Usability (ISFU) system, customized and adapted to our working group needs.

Importance to measure and report	Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high priority aspect of healthcare where there is variation in or overall less-than-optimal performance. Measures must be judged to meet all subcriteria to pass this criterion and be evaluated against the remaining criteria.	**1a. Evidence base** The measure focus is evidence-based: – **Health outcome:** a rationale supports the relationship of the health outcome to processes or structures of care. – A systematic assessment and grading of the quantity, quality, and consistency of the evidence that the measured **structure, process or intermediate clinical outcome** leads to a desired health outcome.**1b. Performance gap** Demonstration of quality problems and opportunity for improvement **1c. High priority** A high priority aspect of healthcare.
Scientific acceptability of measure properties	Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. Measures must be judged to meet the subcriteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.	**2a. Reliability** The measure is well defined and precisely specified so it can be implemented consistently and allows for comparability. **2b. Validity** The measure specifications are consistent with the evidence. Target population and exclusions are supported by the evidence. Validity testing demonstrates that the measure correctly reflects the quality of care provided, adequately identifying differences in quality. Where an evidence-based risk-adjustment strategy is specified, it has demonstrated adequate discrimination and calibration. Analysis of computed measure scores demonstrates that scoring allows for identification of statistically significant and practically/clinically meaningful differences in performance. If multiple data sources/methods are specified, there is demonstration they produce comparable results. For measures susceptible to missing data, analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that results are not biased due to it and how the specified handling of missing data minimizes bias. **2c. Disparities** If disparities in care have been identified, measure specifications, scoring, and analysis allow for identification of disparities through stratification of results.
Feasibility	Extent to which the specifications, including measure logic, required data that are readily available or could be captured without undue burden and can be implemented for performance measurement.	**3a.** For clinical measures, the required data elements are routinely generated and used **3b.** The required data elements are available in electronic sources, or a credible path to electronic collection is specified. **3c.** Demonstration that the data collection strategy can be implemented
Usability and use	Extent to which potential audiences (e. g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high quality, efficient healthcare for individuals or populations.	A credible rationale describes how the performance results could be used to further the goal of high quality, efficient healthcare for individuals or populations.
Comparison to related or competing measures	If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure.	**Consider multiple measures in a domain if:** The measure is harmonized with related measures or multiple measures are justified. **Consider replacing existing measure if:** The measure is superior to existing measures

creases the accuracy of the performance measurement (i. e., it reduces the probability that apparent underperformance is a chance event). Secondly, there is evidence that endoscopy proficiency increases with increasing number of procedures performed, and that endoscopy complications are more common with endoscopists who perform fewer procedures per year [1]; this is also well described in many other clinical areas such as surgery [53]. A trend towards fewer endoscopists each performing more procedures may be appropriate, and setting a minimum number of procedures per year for endoscopists may be one strategy to improve quality.

It is important that we help endoscopists with lower levels of performance to improve. Quality assurance should be about improvement, not punishment. One of the biggest gains in endoscopy quality improvement would be to raise the standards of the lower performers to above minimum quality standard thresholds. Various organizations have developed structured processes for the management of underperforming endoscopists, and experience shows that when handled sensitively but robustly, most endoscopists embrace such support. However, there may at times be barriers to the uptake of endoscopy quality improvement by individuals and even services, ranging from complacen-

Performance measure	[name]
Description	Provide a concise summary statement of performance measure
Domain	[domain name]
Category	Structure/Process/Outcome
Rationale	Explain the importance of the measure
Evidence for performance measure	Use GRADE system for **evidence base** and for **strength of recommendation**
Details	Clearly describe: Target population (denominator) Identification of those from the target population who achieved the specific measure focus (numerator, target condition, event, outcome) Measurement time window Exclusions Risk adjustment/stratification Definitions Data source and feasibility Consider handling of missing data Specifications for composite performance measures include: component measure specifications (unless individually endorsed); aggregation and weighting rules; handling of missing data; standardizing scales across component measures; required sample sizes
Scoring	Describe how the performance measure is calculated (e. g. mean/median, count, ratio, rate/proportion) Indicate if stratification/case mix adjustment or weighting required Frequency of calculation. Describe level of analysis (e. g. individual endoscopist – cecal intubation rate; or service level – bowel preparation quality)
Minimum/target standards	Describe minimum/target standards State "no current standard defined" where none exists Describe how score should be interpreted relative to the minimum/target standard Describe whether the standard includes any tolerance for any factors Describe action that should be taken when performance does not reach minimum standard

Table 6 Customized and adapted descriptive framework for each final performance measure.

cy ("I'm fine and don't need to measure") to fear that one's abilities might be demonstrated to be suboptimal. The latter may be particularly relevant if there are financial or service imperatives to continue with the status quo. Nevertheless, we owe it to our patients to overcome these barriers to ensure that endoscopy is of the highest quality.

ESGE and UEG have identified quality of endoscopy as a major priority. We recommend that all units develop mechanisms for audit and feedback of endoscopist and service performance, using the ESGE performance measures that will be published in future issues of *Endoscopy* over the next year. Regional and national organizations have a responsibility to support and, where required, provide resources for such quality improvement initiatives. We urge all endoscopists and endoscopy services to prioritize quality and to ensure that these performance measures are implemented and monitored at a local level, so that we can provide the highest possible care for our patients.(

Acknowledgments

The authors gratefully acknowledge the contributions from: Stuart Gittens, ECD Solutions in development and running of the web platform; Iwona Escreet and all at Hamilton Services for project administrative support; The Scottish Intercollegiate Guidelines Network, especially Duncan Service, for hosting the critical appraisal module; and The Research Foundation – Flanders (FWO), for funding for Prof. Raf Bisschops.

Competing interests: M. Rutter's department receives research funding from Olympus for a colitis surveillance trial (2014 to present). C. Senore's department receives PillCam Colon devices from Covidien-Given for study conduct, and loaner Fuse systems from EndoChoice. R. Bisschops has received: speaker's fees from Covidien (2009 – 2014) and Fujifilm (2013); speaker's fee and hands-on training sponsorship from Olympus Europe (2013 – 2014); speaker's fee and research support from Pentax Europe; and an editorial fee from Thieme Verlag as co-editor of Endoscopy. R. Valori is a director of Quality Solutions for Healthcare, a company providing consultancy for improving quality and training in healthcare. C. Spada has received training support from Given Imaging (2013 and 2014). M. Bretthauer receives funds from Thieme Verlag for editorial work for Endoscopy. C. Bennett owns and works for Systematic Research Ltd, and received a consultancy fee from ESGE to provide scientific, technical, and methodological expertise for the present project. C. Hassan has received equipment on loan from Fujinon, Olympus, Endochoice, and Medtronic; and consultancy fees from Medtronic, Alpha-Wasserman, Norgine, and EndoChoice. C. Rees's department receives research funding from Olympus Medical, ARC Medical, Aquilant Endoscopy, Almirall, and Cook (from 2010 to the present). M. Dinis-Ribeiro receives funds from Thieme Verlag for editorial work for Endoscopy; his department has received support from Olympus for teaching protocol (from August 2014 to July 2015). T. Ponchon has received: advisory board member's fees from Olympus, Ipsen Pharma, and Boston Scientific (2014 and 2015) and from Cook Medical (2014); speaker's fees from Fujifilm, Ipsen Pharma, and Olympus (2014 and 2015) and from Covidien (2014); training support from Ferring (2014); and research sup-

port from Boston Scientific and Olympus (2014 and 2015). P. Fockens has been receiving consulting support from Olympus, Fujifilm, Covidien, and Creo Medical. L. Aabakken , C. Bellisario, D. Domagk, T. Hucl, M. Kaminski and S. Minozzi, have no competing interests.

Institutions

[1] Department of Gastroenterology, University Hospital of North Tees, Stockton-on-Tees, Cleveland, UK
[2] School of Medicine, Durham University, UK
[3] CPO Piemonte, AOU Città della Salute e della Scienza, Torino, Italy
[4] Gastroenterology Department, University Hospital Leuven, Leuven, Belgium
[5] Department of Medicine I, Josephs-Hospital Warendorf, Academic Teaching Hospital, University of Münster, Warendorf, Germany
[6] Department of Gastroenterology, Gloucestershire Royal Hospital, Gloucester, UK
[7] Department of Gastroenterological Oncology, The Maria Sklodowska-Curie Memorial Cancer Centre and Institute of Oncology, and Medical Center for Postgraduate Education, Warsaw, Poland
[8] Department of Health Management and Health Economics, University of Oslo, Oslo, Norway
[9] Digestive Endoscopy Unit, Catholic University, Rome, Italy
[10] Department of Transplantation Medicine, Oslo University Hospital, Oslo, Norway
[11] K.G. Jebsen Colorectal Cancer Research Centre, University of Oslo, Oslo, Norway
[12] Centre for Technology Enabled Research, Faculty of Health and Life Sciences, Coventry University, Coventry, UK
[13] Nuovo Regina Margherita Hospital, Rome, Italy
[14] Servicio de Gastroenterologia, Instituto Portugues de Oncologia Francisco Gentil, Porto, Portugal
[15] Department of Gastroenterology and Hepatology, Institute for Clinical and Experimental Medicine, Prague, Czech Republic
[16] Department. of Digestive Diseases, Hôpital Edouard Herriot, Lyon, France
[17] Department of Gastroenterology and Hepatology, Academic Medical Center, University of Amsterdam, Amsterdam, the Netherlands

References

1 *Rutter MD*, *Rees CJ*. Quality in gastrointestinal endoscopy. Endoscopy 2014; 46: 526 – 528
2 *Rajasekhar P*, *Rutter M*, *Bramble M* et al. Achieving high quality colonoscopy: Using graphical representation to measure performance and reset standards. Colorectal Dis 2012; 14: 1538 – 1545
3 *Baillie J*, *Testoni PA*. Are we meeting the standards set for ERCP? Gut 2007; 56: 744 – 746
4 *Cotton PB*. Are low-volume ERCPists a problem in the United States? A plea to examine and improve ERCP practice – NOW Gastrointest Endosc 2011; 74: 161 – 166
5 *Williams EJ*, *Taylor S*, *Fairclough P* et al. Risk factors for complication following ERCP; results of a large-scale, prospective multicenter study. Endoscopy 2007; 39: 793 – 801
6 *Williams EJ*, *Taylor S*, *Fairclough P* et al. Are we meeting the standards set for endoscopy? Results of a large-scale prospective survey of endoscopic retrograde cholangio-pancreatograph practice Gut 2007; 56: 821 – 829
7 *Pabby A*, *Schoen RE*, *Weissfeld JL* et al. Analysis of colorectal cancer occurrence during surveillance colonoscopy in the dietary Polyp Prevention Trial. Gastrointest Endosc 2005; 61: 385 – 391
8 *Robertson DJ*, *Lieberman DA*, *Winawer SJ* et al. Colorectal cancers soon after colonoscopy: a pooled multicohort analysis. Gut 2014; 63: 949 – 956
9 *van Rijn JC*, *Reitsma JB*, *Stoker J* et al. Polyp miss rate determined by tandem colonoscopy: a systematic review. Am J Gastroenterol 2006; 101: 343 – 350
10 *Van Gelder RE*, *Nio CY*, *Florie J* et al. Computed tomographic colonography compared with colonoscopy in patients at increased risk for colorectal cancer. Gastroenterology 2004; 127: 41 – 48
11 *Pickhardt PJ*, *Choi JR*, *Hwang I* et al. Computed tomographic virtual colonoscopy to screen for colorectal neoplasia in asymptomatic adults. N Engl J Med 2003; 349: 2191 – 2200
12 *Rockey DC*, *Paulson E*, *Niedzwiecki D* et al. Analysis of air contrast barium enema, computed tomographic colonography, and colonoscopy: prospective comparison. Lancet 2005; 365: 305 – 311
13 *Miller RE*, *Lehman G*. Polypoid colonic lesions undetected by endoscopy. Radiology 1978; 129: 295 – 297
14 *Pickhardt PJ*, *Nugent PA*, *Mysliwiec PA* et al. Location of adenomas missed by optical colonoscopy. Ann Intern Med 2004; 141: 352 – 359
15 *Barclay RL*, *Vicari JJ*, *Doughty AS* et al. Colonoscopic withdrawal times and adenoma detection during screening colonoscopy. N Engl J Med 2006; 355: 2533 – 2541
16 *Chen SC*, *Rex DK*. Endoscopist can be more powerful than age and male gender in predicting adenoma detection at colonoscopy. Am J Gastroenterol 2007; 102: 856 – 861
17 *Pohl H*, *Srivastava A*, *Bensen SP* et al. Incomplete polyp resection during colonoscopy – results of the complete adenoma resection (CARE) study. Gastroenterology 2013; 144: 74 – 80.e1
18 *Singh H*, *Nugent Z*, *Demers AA* et al. The reduction in colorectal cancer mortality after colonoscopy varies by site of the cancer. Gastroenterology 2010; 139: 1128 – 1137
19 *Baxter NN*, *Goldwasser MA*, *Paszat LF* et al. Association of colonoscopy and death from colorectal cancer. Ann Intern Med 2009; 150: 1 – 8
20 *Brenner H*, *Hoffmeister M*, *Arndt V* et al. Protection from right- and left-sided colorectal neoplasms after colonoscopy: population-based study. J Natl Cancer Inst 2010; 102: 89 – 95
21 *Baxter NN*, *Warren JL*, *Barrett MJ* et al. Association between colonoscopy and colorectal cancer mortality in a US cohort according to site of cancer and colonoscopist specialty. J Clin Oncol 2012; 30: 2664 – 2669
22 *Lakoff J*, *Paszat LF*, *Saskin R* et al. Risk of developing proximal versus distal colorectal cancer after a negative colonoscopy: a population-based study. Clin Gastroenterol Hepatol 2008; 6: 1117 – 1121
23 *Singh H*, *Nugent Z*, *Demers AA* et al. Rate and predictors of early/missed colorectal cancers after colonoscopy in Manitoba: a population-based study. Am J Gastroenterol 2010; 105: 2588 – 2596
24 *Brenner H*, *Chang-Claude J*, *Seiler CM* et al. Does a negative screening colonoscopy ever need to be repeated? Gut 2006; 55: 1145 – 1150
25 *Brenner H*, *Chang-Claude J*, *Seiler CM* et al. Protection from colorectal cancer after colonoscopy: a population-based, case-control study. Ann Intern Med 2011; 154: 22 – 30
26 *Valori RM*, *Morris JE*, *Thomas JD* et al. Tu1485 Rates of post colonoscopy colorectal cancer (PCCRC) are significantly affected by methodology, but are nevertheless declining in the English NHS [abstract]. Gastrointest Endosc 2014; 79: AB451 DOI 10.1016/j.gie.2014.02.931
27 *Yalamarthi S*, *Witherspoon P*, *McCole D* et al. Missed diagnoses in patients with upper gastrointestinal cancers. Endoscopy 2004; 36: 874 – 879
28 *Raftopoulos SC*, *Segarajasingam DS*, *Burke V* et al. A cohort study of missed and new cancers after esophagogastroduodenoscopy. Am J Gastroenterol 2010; 105: 1292 – 1297
29 *Cohen J*, *Safdi MA*, *Deal SE* et al. Quality indicators for esophagogastroduodenoscopy. Am J Gastroenterol 2006; 101: 886 – 891
30 *Faigel DO*, *Pike IM*, *Baron TH* et al. Quality indicators for gastrointestinal endoscopic procedures: an introduction. Am J Gastroenterol 2006; 101: 866 – 872
31 *Park WG*, *Cohen J*. Quality measurement and improvement in upper endoscopy. Techniques Gastrointest Endosc 2012; 14: 13 – 20
32 *Gavin DR*, *Valori RM*, *Anderson JT* et al. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. Gut 2013; 62: 242 – 249
33 *Enochsson L*, *Swahn F*, *Arnelo U* et al. Nationwide, population-based data from 11,074 ERCP procedures from the Swedish Registry for Gallstone Surgery and ERCP. Gastrointest Endosc 2010; 72: 1175 – 1184, 1184.e1 – e3
34 *Baron TH*, *Petersen BT*, *Mergener K* et al. Quality indicators for endoscopic retrograde cholangiopancreatography. Am J Gastroenterol 2006; 101: 892 – 897
35 *Cotton PB*, *Garrow DA*, *Gallagher J* et al. Risk factors for complications after ERCP: a multivariate analysis of 11,497 procedures over 12 years. Gastrointest Endosc 2009; 70: 80 – 88
36 Consortium. TANS. Appraisal of guidelines for research and evaluation II. AGREE II Instrument. 2009: 1 – 56
37 *Jaeschke R*, *Guyatt GH*, *Dellinger P* et al. Use of GRADE grid to reach decisions on clinical practice guidelines when consensus is elusive. BMJ 2008; 337: a744 DOI 10.1136/bmj.a744
38 *Murphy MK*, *Black NA*, *Lamping DL* et al. Consensus development methods, and their use in clinical guideline development. Health Technol Assess 1998; 2: i – iv, 1 – 88
39 *Greenhalgh T*. How to read a paper. Getting your bearings (deciding what the paper is about). BMJ 1997; 315: 243 – 246
40 *O'Connor D*, *Green S*, *Higgins JP*. Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S editors.

Cochrane handbook for systematic reviews of interventions. Oxford, UK: Wiley-Blackwell; 2008
41 *Richardson WS*, *Wilson MC*, *Nishikawa J* et al. The well-built clinical question: a key to evidence-based decisions. ACP J Club 1995; 123: A12 – A13
42 *Shea BJ*, *Grimshaw JM*, *Wells GA* et al. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol 2007; 7: 10
43 *Higgins JP*, *Altman DG*, *Gotzsche PC* et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011; 343: d5928
44 *Wells GA*, *Shea B*, *O'Connell DJ* et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. Available at: http://www.ohri.ca/programs/clinical_epidemiology/oxford.htm Accessed: 2015
45 *Whiting PF*, *Rutjes AW*, *Westwood ME* et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011; 155: 529 – 536
46 Effective Practice and Organisation of Care (EPOC). Suggested risk of bias criteria for EPOC reviews. EPOC. Resources for review authors. Oslo.: Norwegian Knowledge Centre for the Health Services; Available at: http://epoc.cochrane.org/epoc-specific-resources-review-authors Accessed: 2015
47 *Guyatt GH*, *Oxman AD*, *Vist GE* et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008; 336: 924 – 926
48 *Guyatt GH*, *Oxman AD*, *Kunz R* et al. What is "quality of evidence" and why is it important to clinicians? BMJ 2008; 336: 995 – 998
49 GRADE Working Group. http://www.gradeworkinggroup.org/ Accessed 2015
50 ISFU system; National Quality Measures Clearinghouse (NQMC). http://www.qualitymeasures.ahrq.gov
51 ISFU criteria; National Quality Measures Clearinghouse (NQMC). http://www.qualitymeasures.ahrq.gov
52 *Gavin DR*, *Valori RM*, *Anderson JT* et al. The national colonoscopy audit: a nationwide assessment of the quality and safety of colonoscopy in the UK. Gut 2013; 62: 242 – 249
53 *Birkmeyer JD*, *Stukel TA*, *Siewers AE* et al. Surgeon volume and operative mortality in the United States. N Engl J Med 2003; 349: 2117 – 2127