

# Latent composite indicators for evaluating adherence to guidelines in patients with a colorectal cancer diagnosis

Rossella Murtas, PhD, Adriano Decarli, PhD, Maria Teresa Greco, MD, Anita Andreano, MD, Antonio Giampiero Russo, MD, PhD\*

## Abstract

Evidence-based guidelines for the correct management of cancer patients are developed on the idea that timely care can improve health prognoses and quality of life.

The aim of this paper is to evaluate the adherence of clinical pathways to clinical guidelines provided at the hospital level, for colorectal cancer care.

By using a retrospective observational study, we proposed a method for associating each patient to a healthcare provider and modeling adherence as a latent construct governed by a set of 10 influential indicators. These indicators measure the adherence to specific guidelines for diagnosis, surgical treatment, chemotherapy, and follow-up. The model used was that of the item response theory (IRT). When evaluating providers, the IRT allows for a comparison of indicators in terms of their discriminating ability and difficulty, and in terms of their adherence to guidelines. The IRT results were compared with non-latent methods: numerator-based weight and denominator-based weight.

A strong degree of coherence of the indicators in measuring adherence, and a high level of overall agreement between latent and non-latent methods were noted. The IRT approach demonstrated similar providers' evaluations between endoscopy and histological assessment indicators. The greatest discriminating ability among providers could be attributed to all diagnostic exams, while the lowest was associated with follow-up endoscopies. The most difficult indicator to achieve was fecal occult blood test, while follow-up imaging was the easiest.

In a decision-making framework, valuable indications can be derived from the use of IRT models rather than weighting methods. Using IRTs, we were able to highlight the principal indicators in terms of strength of discrimination, and to isolate those that merely duplicated information.

**Abbreviations:** ATS = Agency for Health Protection, CEA = carcinoembryonic antigen, CRC = colorectal cancer, DBW = denominator-based weights, FOBT = fecal occult blood test, ICC = item characteristic curve, IRT = item response theory, NBW = numerator-based weights.

**Keywords:** adherence to guidelines, colorectal cancer, composite indicators, item response theory, quality in health care

## 1. Introduction

Colorectal cancer (CRC) represents the third most commonly diagnosed cancer in both genders<sup>[1]</sup> and is one of the leading

causes of deaths worldwide. For the Italian population, it was ranked in the 10 most significant causes of death for the year 2017, and represents the second most common cause of death from cancer.<sup>[2]</sup> Previous studies have shown that survival rates are influenced by pathways of care that adhere to evidence-based guidelines that have been reported for CRC and other cancers.<sup>[3–8]</sup> However, a large amount of heterogeneity is present across, and within, countries in the definition of the optimal pathway of care,<sup>[9]</sup> as different health care agencies provide different guidelines for caring for CRC patients. The National Institute for Health and Care Excellence (NICE, The United Kingdom) has implemented guidelines for CRC diagnosis, treatment, and follow-up since 2007, which is mainly focused on procedures that have been shown to improve either survival rates or quality of life.<sup>[10]</sup> At the national level, the Italian Association of Medical Oncology provides clinical recommendations for the management of patients with CRC.<sup>[11]</sup> To evaluate adherence to evidence-based guidelines, the current goal is to define sets of indicators, usually defined by teams of experts, which include measures of the quality of care delivered to cancer patients from diagnosis through follow-up.<sup>[12,13]</sup> In Italy, a set of 26 indicators have been constructed from a literature review, with the aim of evaluating the adherence to the entire pathway of treatment for CRC patients, with measures relating to diagnosis, surgery, medical treatment, and follow-up.<sup>[14]</sup>

Editor: Chun Gao.

The Italian Ministry of Health [RF-2011-02348959] supported this work.

The authors have no conflicts of interest to disclose.

Supplemental Digital Content is available for this article.

Epidemiology Unit, Agency for Health Protection of Milan, Corso Italia 19, 20122 Milan, Italy.

\* Correspondence: Antonio Giampiero Russo, Epidemiology Unit, Agency for Health Protection of Milan, Corso Italia 19, 20122 Milan, Italy (e-mail: agrusso@ats-milano.it).

Copyright © 2020 the Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial License 4.0 (CCBY-NC), where it is permissible to download, share, remix, transform, and buildup the work provided it is properly cited. The work cannot be used commercially without permission from the journal.

How to cite this article: Murtas R, Decarli A, Greco MT, Andreano A, Russo AG. Latent composite indicators for evaluating adherence to guidelines in patients with a colorectal cancer diagnosis. *Medicine* 2020;99:8(e19277).

Received: 2 October 2018 / Received in final form: 9 September 2019 /

Accepted: 22 January 2020

<http://dx.doi.org/10.1097/MD.00000000000019277>

Although each indicator describes a different and valuable portion of the pathways of care, there is a compelling need to develop composite measures to evaluate the degree of adherence on a global level.<sup>[10,11]</sup> Several methods have been proposed in the statistical and economic literature to synthesize the indicators. This distinction stems from the use of latent and non-latent methods to define composite scores. Non-latent methods,<sup>[15–17]</sup> such as all-or-none scores, are a weighted average of the individual indicators, which can be performed with either numerator-based weights (NBW) or denominator-based weights (DBW).<sup>[18–20]</sup> On the other hand, latent methods suppose the existence of an unmeasurable variable – the “adherence to guidelines” – which determines the distribution of the indicators. The probability of achieving each indicator can be modeled using the item response theory (IRT) method.<sup>[21,22]</sup> These models have been widely implemented before, especially for Pay-for-Performance programs,<sup>[23]</sup> with the intention of rewarding hospitals that have a superior level of performance. Other applications can be found in the evaluation of care delivered in cases of acute myocardial infarction, congestive heart failure and community-acquired pneumonia.<sup>[16,24,25]</sup> However, it is not equally common to find studies using the IRT method in the field of oncology.

The aim of our study was to evaluate the levels of adherence of CR pathways to the clinical guidelines provided at the hospital level, by comparing latent and non-latent methods, and to assess how each indicator discriminates against providers across various levels of adherence.

## 2. Methods

A retrospective cohort study was performed, including all new diagnoses of the colon and rectal cancer, occurring from 2007 to

2012, excluding non-epithelial cancers (i.e., cases of lymphoma and sarcoma), and cases identified by death certificate only (DCO), in the population of the Milan municipality. This geographical area is covered and accredited by the International Agency for Research on Cancer. It is further included in the Cancer Incidence in Five Continents – Volume XI,<sup>[26]</sup> which continuously collects data on all new invasive cancers and covers the entire province of Milan with 3,480,513 inhabitants.

The study was carried out in accordance with the principles established in the Declaration of Helsinki. The study was not reviewed by the Ethics Committee, because it is a retrospective observational study based on cases routinely collected by the cancer registry of the province of Milan. The Agency for Health Protection (ATS) of the province of Milan is authorized by the Guarantor of Privacy to use this health-related data for analysis purposes. Anonymization was ensured by means of an internal code that is used in every administrative database belonging to the ATS of Milan.

In a recent project, 26 indicators have been constructed through a review of the literature and the use of previously described guidelines.<sup>[14]</sup> Each indicator measures adherence to a recommended procedure for CRC patients. The indicators were calculated using specific algorithms, starting from CRC cases identified by the population-based cancer registry of Milan (Italy),<sup>[26]</sup> which is linked with the healthcare databases. The indicators cover the entire pathway of treatment for CRC patients, from diagnosis (eight indicators), through surgery (seven indicators), medical treatment (seven indicators), and follow-up (four indicators). From this set of indicators, we selected the 10 measures with the highest level of correlation, according to Cronbach alpha as an estimate of internal reliability.<sup>[27]</sup> The measures included are described in Table 1,

**Table 1**  
Selected indicators.

Indicator	Name	Description
D <sub>1</sub>	FOBT	<b>Definition:</b> Proportion of patients receiving FOBT within 6 months before and 1 month after diagnosis. <b>Eligible patients:</b> IF FOBT from screening, patients aged 49 to 69.
D <sub>2</sub>	Endoscopy	<b>Definition:</b> Proportion of patients undergoing endoscopy within 6 months before and 1 month after diagnosis. Patients hospitalized 3 months before diagnosis are supposed of receiving endoscopy. <b>Eligible patients:</b> Those who did not received endoscopy because peritonitis or bowel obstruction within 30 days after or before diagnosis.
D <sub>3</sub>	CEA	<b>Definition:</b> Proportion of patients receiving CEA evaluation within 6 months before and 1 month after diagnosis. Patients hospitalized 3 months before diagnosis are supposed of receiving CEA evaluation. <b>Eligible patients:</b> All.
D <sub>4</sub>	Diagnostic imaging	<b>Definition:</b> Proportion of patients undergoing diagnostic imaging within 6 months before and 1 month after diagnosis. <b>Eligible patients:</b> All.
D <sub>5</sub>	All exams at diagnosis	<b>Definition:</b> Proportion of patients undergoing endoscopy, CEA evaluation and diagnostic imaging within 6 months before and 1 month after diagnosis. <b>Eligible patients:</b> Those who did not received endoscopy because peritonitis or bowel obstruction within 30 days after or before diagnosis.
D <sub>6</sub>	Histological assessment	<b>Definition:</b> Proportion of patients (who received a surgical treatment) undergoing a histological assessment within 3 months before surgery. Patients hospitalized 3 months before diagnosis are supposed of receiving the treatment. <b>Eligible patients:</b> Those who did not receive the procedure because peritonitis or bowel obstruction within 30 days after or before diagnosis and who received a surgical treatment.
D <sub>7</sub>	Follow-up endoscopy	<b>Definition:</b> Proportion of patients undergoing endoscopy within 12 months after surgery. <b>Eligible patients:</b> Patients not in stage IV, who received a surgical treatment.
D <sub>8</sub>	Follow-up CEA	<b>Definition:</b> Proportion of patients undergoing CEA evaluation within 12 months after surgery. <b>Eligible patients:</b> Patients not in stage IV, who received a surgical treatment.
D <sub>9</sub>	Follow-up imaging	<b>Definition:</b> Proportion of patients undergoing diagnostic imaging within 12 months after surgery. <b>Eligible patients:</b> Patients not in stage IV, who received a surgical treatment.
D <sub>10</sub>	All follow-up exams	<b>Definition:</b> Proportion of patients undergoing endoscopy, CEA evaluation and diagnostic imaging within 12 months after surgery. <b>Eligible patients:</b> Patients not in stage IV, who received a surgical treatment.

FOBT = fecal occult blood test, CEA = carcinoembryonic antigen.

together with a brief definition of the individual characteristics required for eligibility. For example, patients with peritonitis or bowel obstruction close to the date of diagnosis (within 30 days before or after diagnosis) were deemed ineligible for the indicator that evaluates the execution of pre-diagnosis endoscopies (D<sub>2</sub>). Adherence to guidelines is also captured by the time frame in which each the indicator was evaluated. For example, for diagnosis indicators (D<sub>1</sub>–D<sub>5</sub>), a patient was considered to be compliant if he/she received the diagnostic procedure within 6 months before, and 1 month after, the diagnosis. As a second illustration, we know that guidelines require a histological determination for each CRC patient receiving surgical treatment. For this reason, the indicator D<sub>6</sub> identified the proportion of patients undergoing surgery, with a histological assessment in the 3 months prior to the surgical intervention. Finally, for the follow-up indicators (D<sub>7</sub>–D<sub>10</sub>), a period of 1 year after surgery was permitted.

### 2.1. Provider accountability

The assignment of a patient to a single provider – which is considered to be the principal subject responsible for the care received – is a very difficult step because individual pathways of care are not uniquely associated with the same hospital. Chen et al.<sup>[23]</sup> addressed this problem by applying the *plurality provider algorithm*, which assigns a patient to the provider who billed the greatest number of care visits in a given year. For CRC patients, the fundamental step in the care process is represented by primary surgery. For this reason, in order to evaluate adherence to guidelines at the provider level, we associated each patient with the hospital where the first major surgical treatment was performed. If no operation had been performed, we associated the patient to the care provider where the endoscopy, chemotherapy or carcinoembryonic antigen (CEA) dosage were performed, in this order of priority. Patients who did not undergo any of these procedures were associated with the hospital where they were first admitted with a CRC diagnosis (ICD9 codes: 153–154).<sup>[3]</sup> It is clear that different allocation methods will dictate different distributions of adherence across providers. However, this method ensured that, in cases where no surgery had been performed, the patient was associated with the provider where the first step of care was administered. Given that a healthcare organization can be composed of more than one structure or provider, we decided to perform the analysis at the provider level. However, the results are displayed as the estimated average adherence across providers that belong to the same organization. Furthermore, considering that the aim of the study was to compare organizations within the local health authority (ATS) of Milan, we chose to plot adherence levels only for those organizations that belong to the area governed by said ATS.

## 2.2. Statistical methods

### 2.2.1. Indicators of adherence for selected procedures.

For each indicator, and for each provider, we have defined  $n_{ij}$  as the number of eligible patients for the  $i^{th}$ -procedure, which had been administered by the  $j^{th}$ -provider, and  $y_{ij}$  as the number of patients assigned to the  $j^{th}$ -provider who underwent procedure  $i$  (even if it was not administered by the  $j^{th}$  provider). For example, according to the reference guidelines, each patient should have a histological characterization of the tumor before their surgery. Therefore, we considered a patient with a biopsy record in the 3 months prior to the date of the surgery as being in compliance with this

recommendation, regardless of whether it was performed by the same provider or not. The indicator measuring the recommended procedure,  $i$ , hereafter referred to as indicator  $i$ , for the  $j^{th}$ -provider is calculated as the proportion  $D_{ij} = y_{ij}/n_{ij}$ .

**2.2.2. Composite indicators: non-latent methods.** The DBW method assigns a weight to each of the 10 individual indicators for the proportion of eligible patients in the population:

$$DBW_{.j} = \frac{n_{1j}}{\sum_{i=1}^{10} n_{ij}} \times \frac{y_{1j}}{n_{1j}} + \dots + \frac{n_{10j}}{\sum_{i=1}^{10} n_{ij}} \times \frac{y_{10j}}{n_{10j}} = \frac{\sum_{i=1}^{10} y_{ij}}{\sum_{i=1}^{10} n_{ij}}. \quad (1)$$

On the other hand, the NBW method assigns a weight to each of the individual indicators for the proportion of patients receiving the associated procedure:

$$NBW_{.j} = \frac{y_{1j}}{\sum_{i=1}^{10} y_{ij}} \times \frac{y_{1j}}{n_{1j}} + \dots + \frac{y_{10j}}{\sum_{i=1}^{10} y_{ij}} \times \frac{y_{10j}}{n_{10j}}. \quad (2)$$

Providers with high levels of DBW or NBW will be considered to be highly adherent to guidelines.

The choice between DBW and NBW should depend on the aim of the analysis, and there are criticisms of both approaches. The former gives more weight to those indicators with a higher proportion of eligibility, while the latter assigns more weight to the indicators that have a higher proportion of compliance. According to Babbie,<sup>[28]</sup> items should be equally weighted unless there are compelling reasons for differentiated weighting. If indicators share the same number of eligible patients across providers, then the DBW method is found to be most in keeping with this requirement. However, these methods do not deal with the correlation that exists between indicators, and their treatment of missing values is insufficient. In addition, they have been widely used to investigate constructs in formative models, where adherence is intended to be caused by the indicators themselves.<sup>[20,29]</sup>

**2.2.3. Composite indicators: latent methods.** In order to synthesize the information resulting from the 10 selected indicators, we used the latent IRT method. This model has primarily been employed in the fields of psychology and education.<sup>[21,22,30]</sup> It has been used to measure abilities, by administering questionnaires composed of several questions – called *items* – with binary, polytomous, or categorical answers.<sup>[23,31]</sup> Latent models assume the existence of an unmeasurable trait (e.g., the propensity to adherence), which is somehow connected to the observed variables (e.g., the indicators). IRTs assume a reflective construct, where items are caused by the latent variable, such that they are assumed to be independent.

This model attempts to measure adherence to guidelines  $\theta_i$  (for  $j=1, \dots, N$ ) using the binomial variable  $y_{ij}$ , with parameters  $n_{ij}$  and  $p_{ij}$ , such that:

$$\text{probit}(p_{ij}) = \alpha_i(\theta_i - \beta_i). \quad (3)$$

where  $p_{ij}$  is the probability of a patient receiving procedure  $i$  administered by provider  $j$ . Here, we assume that  $\theta$  is normally distributed, with a mean of 0 and a variance of 1, that is, higher values of  $\theta$  will describe more adherent providers. The model in Eq. (3) is referred to as the two-parameter Normal-Ogive model,<sup>[21]</sup> where the latent trait is connected to the probability of receiving procedure  $i$  via two parameters,  $\alpha_i$  and  $\beta_i$ . The parameter  $\alpha_i$  is referred to as the discrimination parameter and represents the strength of the relationship between an indicator

and the latent level of adherence of the  $j^{th}$  provider. The parameter  $\beta_i$  is referred to as the difficulty parameter and represents the median of  $\theta$  for the  $i^{th}$  indicator. It has been proven elsewhere that there is a correspondence between the difficulty parameter,  $\beta_i$ , and the indicator's proportion of correct responses.<sup>[32]</sup>

For each indicator  $i$ , we can define an item characteristic curve (ICC), which is roughly equivalent to the regression of  $y_{ij}$  (on the probit scale) on the distribution of adherence levels  $\theta$ . The shape of each ICC is governed by the discrimination  $\alpha$ , that is, the higher the value, the more the indicator discriminates against the adherence to guidelines of the underlying provider. The difficulty parameter shifts each curve on the horizontal axis, that is, it represents the minimum level of adherence for which the probability of receiving the procedure is higher than 0.5. In order to describe all providers, ranging from the least to the most adherent, a good set of indicators should cover the entire range of  $\theta$ .

Providers with no eligible patients for the  $i^{th}$ -indicator (and, consequently, zero patients who are undergoing procedure  $i$ ) were considered missing for the adherence evaluation for that indicator. Missing indicators are not treated insufficiently in IRT models. In fact, the IRT method is the perfect tool for handling data with missing values, as missing responses do not contribute to the evaluation of adherence.

IRT models presuppose different hypotheses for identifiability, such as monotonicity on the ICC, local independence among indicators, and the uni-dimensionality of the latent trait. The hypothesis of uni-dimensionality was tested using Cronbach alpha as an estimate of internal reliability.<sup>[27]</sup> The choice of the selected set of indicators was based on a minimum item-total correlation value of 0.3.<sup>[33]</sup> Higher positive values indicate the appropriateness of the item for the latent construct, that is, adherence to guidelines. The adequacy of the sample was evaluated using Kaiser measure.<sup>[34]</sup> Values between 0.8 and 0.9 are considered to be adequate, while values below 0.5 are considered unacceptable.

The estimated adherence level was plotted using funnel plots, where the number of patients assigned to the corresponding provider was represented on the x-axis. Finally, we compared provider adherence levels, which were estimated using latent and non-latent methods, by means of Spearman rank correlation. All the analyses were implemented using PROC NL MIXED SAS/

STAT software, Version 9.4, SAS Institute Inc., Cary, NC (see *Supplemental Material*, <http://links.lww.com/MD/D847>, which illustrates IRT implementation in SAS).

### 3. Results

#### 3.1. Description of the data

The cohort included 10,552 subjects with CRC incident in the period 2007 to 2012. We excluded 33 cases of lymphoma/sarcoma and 135 cases of DCO. We further excluded 710 patients that we could not associate with any provider from the analysis, as they had not undergone any of the following procedures – surgical treatment, endoscopy, chemotherapy or CEA dosage– and were never admitted to a hospital with a CRC diagnosis. The resulting cohort consisted of 9674 subjects (92% of the CRC patients included). The characteristics of the CRC patients in the original and the restricted cohorts were similar. The 710 people not included in the final analysis were mostly female (60%) and between the ages of 80 and 90 years (30%).

#### 3.2. Provider accountability

In accordance with the provider accountability method, 79% of the cases were associated with the hospital where they underwent their first surgical treatment, 13.4% with the provider where they underwent endoscopy, 3.3% with where they received chemotherapy, 2% with where they underwent CEA evaluation, and 2.4% with the hospital where their first admission with a CRC diagnosis occurred. The algorithm identified 135 providers, which were grouped into 68 organizations. Of them, 42 were in the Milan area and 26 were outside of it. Among those 42 organizations, 16 were laboratories associated with CEA evaluation. We presented the estimated adherence results for the remaining 26 organizations.

#### 3.3. Indicators of adherence for selected procedures

Table 2 displays descriptive statistics for each of the selected indicators. Cronbach alpha and the item-total correlation were calculated for the subset of 89 providers without missing values. Cronbach alpha was 0.82, indicating a high level of reliability

**Table 2**

**Descriptive statistics of the 10 selected indicators and item response theory model parameters of difficulty and discrimination for each indicator.**

Indicator	Name	N*	Min	Median	Max	Item-total correlation <sup>†</sup>	Difficulty parameter	Discrimination parameter
							$\hat{\beta}$ (Pr >  t )	$\hat{\alpha}$ (Pr >  t )
D <sub>1</sub>	FOBT	135	0	0.18	1	0.49	2.91 (<.01)	0.24 (<.01)
D <sub>2</sub>	Endoscopy	132	0	0.59	1	0.65	-0.86 (<.01)	0.49 (<.01)
D <sub>3</sub>	CEA	135	0	0.50	1	0.47	0.44 (<.01)	0.56 (<.01)
D <sub>4</sub>	Diagnostic imaging	135	0	0.72	1	0.44	-0.86 (<.01)	0.43 (<.01)
D <sub>5</sub>	All exams at diagnosis	132	0	0.21	1	0.64	0.67 (<.01)	0.62 (<.01)
D <sub>6</sub>	Histological assessment	92	0	0.77	1	0.49	-0.83 (<.01)	0.58 (<.01)
D <sub>7</sub>	Follow-up endoscopy	91	0	0.40	1	0.32	1.76 (<.01)	0.22 (<.01)
D <sub>8</sub>	Follow-up CEA	91	0	0.73	1	0.50	-2.32 (<.01)	0.25 (<.01)
D <sub>9</sub>	Follow-up imaging	91	0	0.86	1	0.49	-2.69 (<.01)	0.32 (<.01)
D <sub>10</sub>	All follow-up exams	91	0	0.33	1	0.53	2.36 (<.01)	0.23 (<.01)

FOBT = fecal occult blood test, CEA = carcinoembryonic antigen.

\* Number of providers without missing values for the correspondent indicator.

† Calculated on the subset of 89 providers without missing values.

**Table 3**  
**Descriptive statistics and Spearman correlation among latent and non-latent composites indicators.**

	N*	$\theta^\dagger$ (IRT)	Min	Median	Media	S.D.	Max
DBW	89	0.98	0.20	0.55	0.54	0.16	1
NBW	89	0.59	0.31	0.66	0.69	0.16	1
$\theta$ (IRT)	135	1	-2.04	0.31	0.21	0.91	2.62

DBW = denominator-based weight, NBW = numerator-based weight, IRT = item response theory.  
 \* Number of providers without missing values for the correspondent composite indicator.  
 † Spearman correlations among composites calculated on the 89 providers without missing values.

among indicators. All of the indicators included in the final analysis had an item-total correlation that was greater than 0.3, indicating a high level of overall appropriateness. Kaiser measure of sampling adequacy, based on the 89 providers without missing values was 0.61.

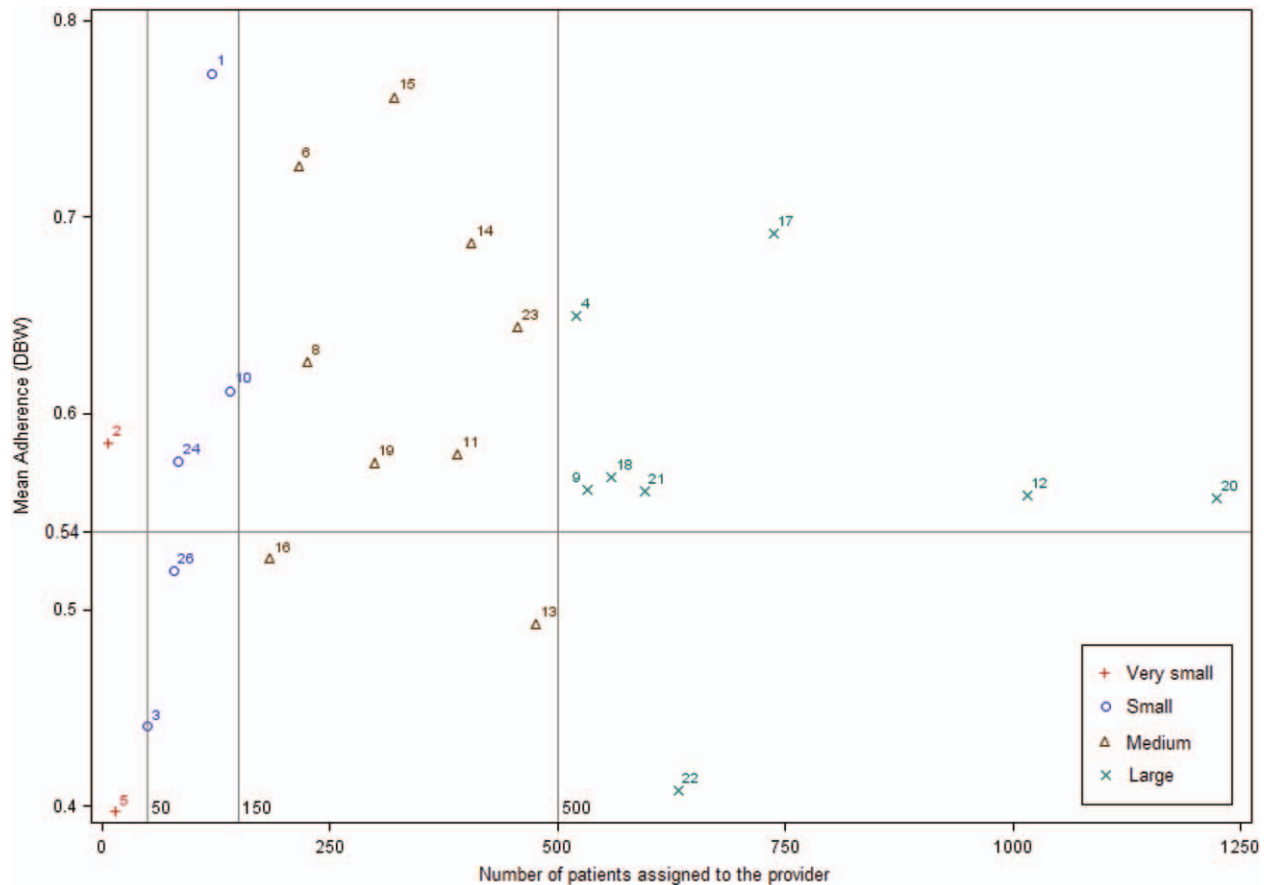
**3.4. Composite indicators**

Descriptive statistics for each composite indicator are shown in Table 3. For each provider, both the DBW and the NBW were between zero and one. However, the DBW displayed a wider range of adherence values. Non-latent composite indicators were calculated for the set of 89 providers without missing values in any of the selected indicators, while adherence levels by IRT were

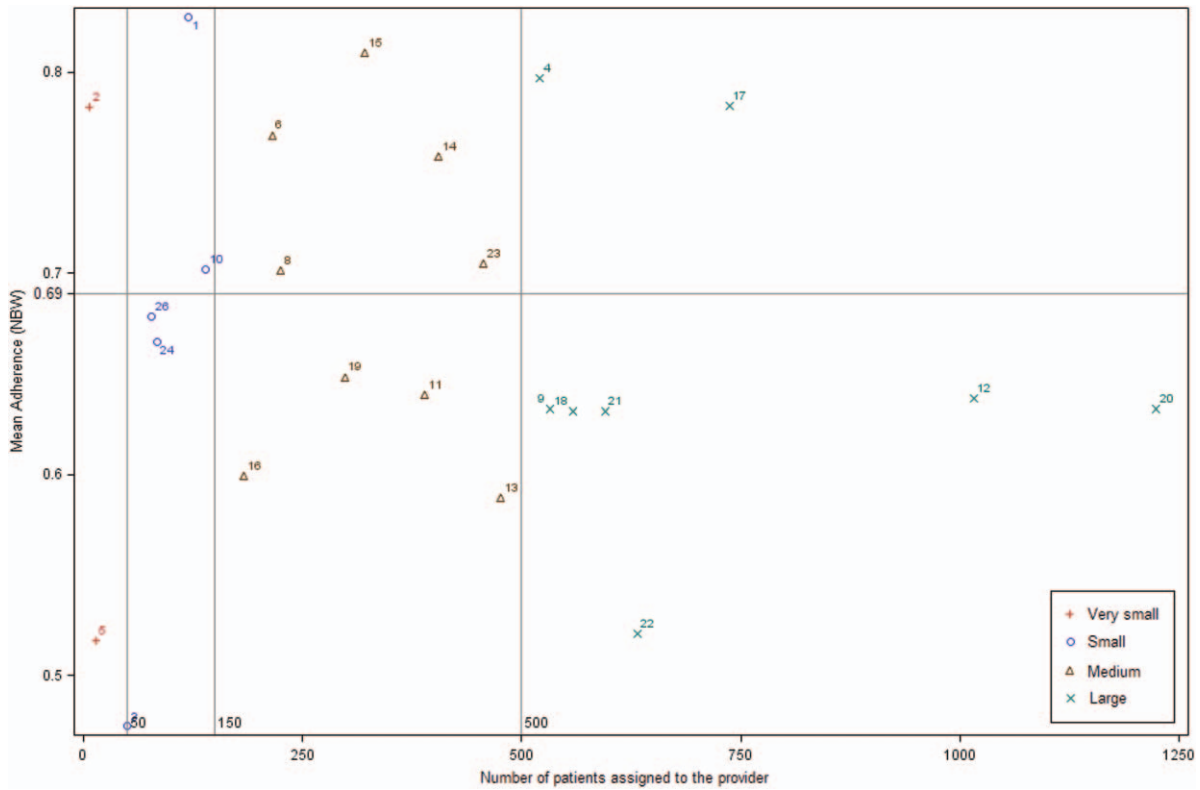
calculated for the entire set of providers. Table 3 shows the Spearman correlation of provider rankings, based on latent and non-latent methods, for the set of 89 providers without missing values. The DBW showed a high level of correlation with the IRT, while the correlation was lower between the IRT and the NBW.

Figures 1–3 show the scatter plots of the distribution of mean adherence, calculated using the three methods (DBW, NBW, and IRT, respectively) and based on the total number of patients assigned to each organization. The organizations were classified as very small (<50 assigned patients), small (between 50 and 150 assigned patients), medium (between 150 and 500 assigned patients), and large (more than 500 assigned patients).

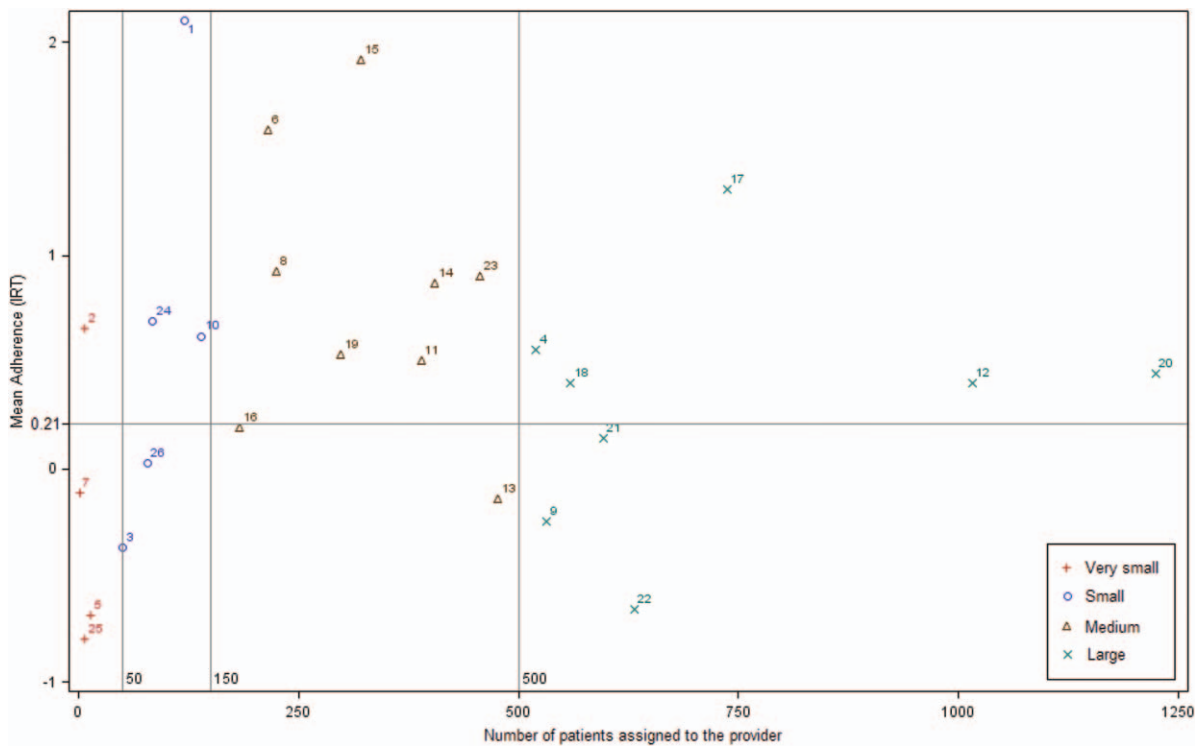
The distribution of adherence across providers was practically identical between the two non-latent methods, with limited



**Figure 1.** Scatter plot of the mean adherence of the 26 (non-laboratory) organizations in the Milan area, based on the total number of patients assigned to each organization. These are divided into very small (<50 assigned patients), small (between 50 and 150 assigned patients), medium (between 150 and 500 assigned patients), and large (more than 500 assigned patients), according to DBW. DBW = denominator-based weights.



**Figure 2.** Scatter plot of the mean adherence of the 26 (non-laboratory) organizations in the Milan area, based on the total number of patients assigned to each organization. These are divided into very small (<50 assigned patients), small (between 50 and 150 assigned patients), medium (between 150 and 500 assigned patients), and large (more than 500 assigned patients), according to NBW. NBW= numerator-based weights.



**Figure 3.** Scatter plot of the mean adherence of the 26 (non-laboratory) organizations in the Milan area, based on the total number of patients assigned to each organization. These are divided into very small (<50 assigned patients), small (between 50 and 150 assigned patients), medium (between 150 and 500 assigned patients), and large (more than 500 assigned patients), according to the IRT method. IRT= item response theory.

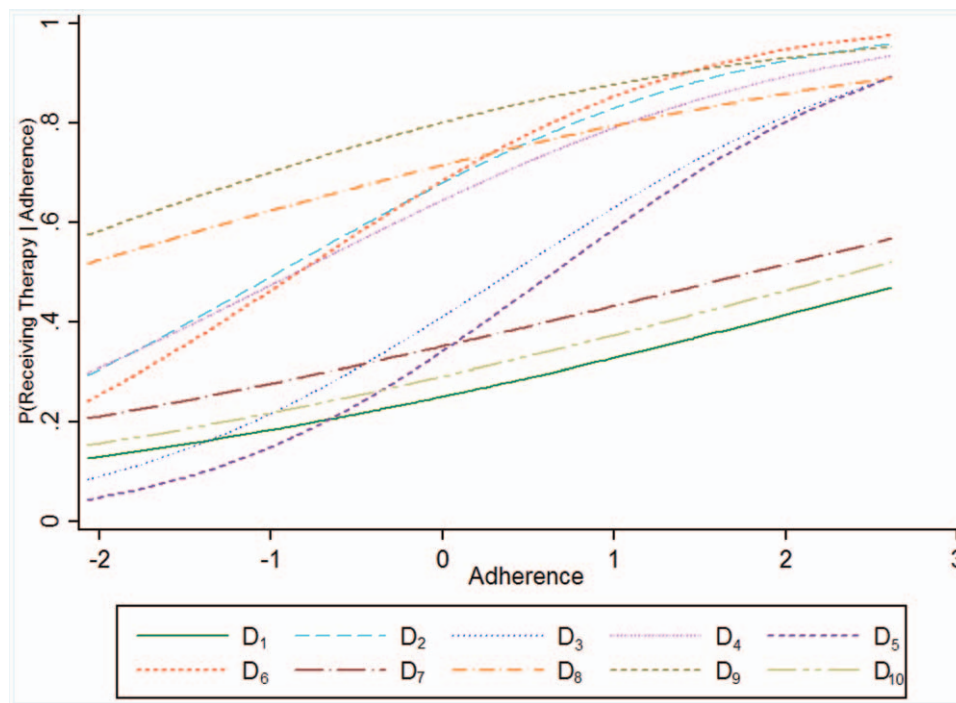


Figure 4. ICC of the 10 selected indicators ( $D_1$ – $D_{10}$ , as described in Table 1). ICC= item characteristic curve.

variation in terms of mean adherence. A slightly different distribution was found with IRT. However, for each of the four categories of assigned patients, the uppermost and lowermost adherent organizations were fairly consistent between the three methods, that is, the most and least adherent organizations were almost identical. However, non-latent methods treat missing values in an insufficient manner. In fact, very small organizations, such as seven and 25, were missing for DBW and NBW, while their adherence could be evaluated using the IRT method. Table 2 shows the difficulty and discrimination parameters for the 10 indicators. The most difficult indicator to achieve was the fecal occult blood test (FOBT,  $D_1$ ), with  $\beta_1=2.91$ , followed by all follow-up exams ( $D_{10}$ ), with  $\beta_{10}=2.36$ . The easiest to achieve was follow-up imaging ( $D_9$ ), with  $\beta_9=-2.69$ . This is consistent with the information given in Table 2: that there is a correspondence between the difficulty parameters and the median of each indicator. The indicator evaluating all diagnostic exams ( $D_5$ ) had the highest degree of discrimination ( $\alpha_5=0.62$ ), while the follow-up endoscopy ( $D_7$ ) had the lowest ( $\alpha_7=0.22$ ). Figure 4 shows the ICCs for the 10 selected indicators. It can be seen that, for Endoscopy ( $D_2$ ) and Histological assessment ( $D_6$ ), the shape describing the probability of receiving the procedure is almost identical. This is in spite of the fact that Histological assessment ( $D_6$ ) had a stronger discrimination parameter. In fact, these indicators have a high level of correlation by definition, given that histological assessment usually results in an endoscopic biopsy. Moreover, the ICC curves for follow-up endoscopy ( $D_7$ ) and all follow-up exams ( $D_{10}$ ) appeared to be quite similar in terms of their discrimination parameters. This is because endoscopy is the most difficult follow-up exam to achieve ( $\beta_7=1.76$  vs  $\beta_8=-2.32$  and  $\beta_9=-2.69$ ), and hence, the majority of people undergoing all follow-up exams ( $D_{10}$ ) had also received a follow-up endoscopy ( $D_7$ ).

#### 4. Discussion

Our study presents a new methodological approach for evaluating the adherence of care at the individual level, but also among hospitals. Indicators evaluating a single recommendation are valid estimates of the adherence of care and each of them should be evaluated on an individual basis. However, within a decision-making framework, we often need composite measures to evaluate the degree of adherence among providers on a global level.

The use of composite indicators in the field of oncology appeared quite recently in the literature, and this can be considered to be one of the first works in this field. Chien et al<sup>[35]</sup> used a simpler version of IRT<sup>[36]</sup> to evaluate the quality of care for CRC patients, using 13 indicators that evaluated both surgical treatment (including pre- and post-assessments) and the colonoscopy before and after surgery. They presupposed equal levels of discrimination between indicators, while we hypothesized that each of these may have a different level of discrimination against the latent level of adherence. Furthermore, they measured the quality of care at the individual level, while we compared adherence at the provider level.

In this study, we used data from a population-based cancer registry to examine differences in the adherence of care among providers, by using indicators based on validated algorithms, established on the basis of healthcare data.<sup>[14]</sup>

However, this method posed some difficult decisions/questions, for example, how to assign providers to patients who underwent procedures in multiple hospitals. We used the provider accountability method to address this issue. In our opinion, the methods available in the literature (such as the *plurality provider algorithm*)<sup>[23,37]</sup> are not appropriate for evaluating the administration of care to CRC patients, because they allocate the patient to the provider who handled the greatest

number of visits, without considering the impact of each procedure on survival. Instead, we chose to adopt a method that closely resembled the typical management of every cancer patient. Another limitation could be found in the decision to use collapsing data about providers. A possible improvement could be obtained by defining a multilevel model for patients' responses to each indicator, considering subjects clustered on providers.

We proposed an allocation method that associated each patient to the provider where, if the surgery had not been performed, the first step of care was administered. We further evaluated overall adherence via two non-latent methods and one latent method. The results showed a high level of overall agreement between the methods, especially between IRT and DBW, as previously observed.<sup>[19]</sup> However, the IRT model is particularly interesting within a decision-making framework. It allows us to distinguish between higher-quality providers by choosing only the procedures with the highest capacity to discriminate, not including indicators with overlapping distribution across providers. For example, endoscopy and histological assessment evaluated providers in a similar manner. Hence, only one of them can and should be chosen to evaluate adherence to guidelines. A greater capacity to discriminate among providers could be attributed to CEA and all of the diagnostic exam indicators, followed by endoscopy, imaging, and histological assessment. The indicators regarding FOBT, follow-up endoscopy and all follow-up exams are characterized by a low level of adherence (<40% in the various providers), while follow-up CEA and follow-up imaging are characterized by a level of adherence that is consistently higher than 50%. According to the IRT method, only seven (out of 26) organizations had a negative level of mean adherence to guidelines (two providers, the 7<sup>th</sup> and 25<sup>th</sup>, have missing information for non-latent methods).

The distribution of the adherence across providers was quite similar, that is, we found a correlation greater than, or equal to, 0.6 between the IRT and DWB methods and the IRT and NBW methods. Greater differences were found in the providers closest to the mean level of adherence. Providers with the highest and lowest adherence values remained relatively consistent, except for the most adherent providers according to the NBW, but not the DBW and the IRT. Thus, we suggest adding a composite score, with IRT models being the preferred method, to globally evaluate providers in terms of their delivery of care that adheres to guidelines for CRC patients.

Finally, the results of this study require confirmation through further application, in order to validate the use of IRT methods in measuring adherence to guidelines. Furthermore, it would be helpful to verify its potential application, not only for other cancer types (breast and lung cancer), but also for non-cancerous, chronic diseases (cardiovascular, neurological and respiratory diseases and diabetes).

## Author contributions

**Conceptualization:** Rossella Murtas, Adriano Decarli, Maria Teresa Greco, Anita Andreano, Antonio Giampiero Russo.

**Data curation:** Rossella Murtas, Maria Teresa Greco.

**Formal analysis:** Rossella Murtas, Maria Teresa Greco.

**Funding acquisition:** Antonio Giampiero Russo.

**Investigation:** Rossella Murtas, Antonio Giampiero Russo.

**Methodology:** Rossella Murtas, Adriano Decarli, Maria Teresa Greco, Anita Andreano, Antonio Giampiero Russo.

**Project administration:** Adriano Decarli, Antonio Giampiero Russo.

**Resources:** Antonio Giampiero Russo.

**Software:** Rossella Murtas, Maria Teresa Greco, Anita Andreano, Antonio Giampiero Russo.

**Supervision:** Adriano Decarli, Maria Teresa Greco, Anita Andreano, Antonio Giampiero Russo.

**Validation:** Rossella Murtas, Adriano Decarli.

**Visualization:** Rossella Murtas.

**Writing – original draft:** Rossella Murtas, Antonio Giampiero Russo.

**Writing – review & editing:** Rossella Murtas, Adriano Decarli, Maria Teresa Greco, Anita Andreano, Antonio Giampiero Russo.

Rossella Murtas orcid: 0000-0002-8989-7605.

## References

- [1] Globocan. 2012. Available at: <http://globocan.iarc.fr/Default.aspx> [accessed July 25, 2018].
- [2] GBD Compare. Institute for Health Metrics and Evaluation. 2014. Available at: <http://www.healthdata.org/data-visualization/gbd-compare> [accessed August 2, 2019].
- [3] Rhoads KF, Ngo JV, Ma Y, et al. Do hospitals that serve a high percentage of medicaid patients perform well on evidence-based guidelines for colon cancer care? *J Health Care Poor Underserved* 2013;24:1180–93.
- [4] Romaguera D, Ward H, Wark PA, et al. Pre-diagnostic concordance with the WCRF/AICR guidelines and survival in European colorectal cancer patients: a cohort study. *BMC Med* 2015;13:107.
- [5] Showalter TN, Camacho F, Cantrell LA, et al. Determinants of quality care and mortality for patients with locally advanced cervical cancer in Virginia. *Medicine (Baltimore)* 2016;95:e2913.
- [6] Søreide K, Traeland J, Stokkeland P, et al. Adherence to national guidelines for surveillance after curative resection of nonmetastatic colon and rectum cancer: a survey among Norwegian gastrointestinal surgeons. *Colorectal Dis* 2012;14:320–4.
- [7] Yen TW, Pezzin LE, Li J, et al. Effect of hospital volume on processes of breast cancer care: a National Cancer Data Base study. *Cancer* 2017;123:957–66.
- [8] Andreano A, Rebora P, Valsecchi MG, et al. Adherence to guidelines and breast cancer patients survival: a population-based cohort study analyzed with a causal inference approach. *Breast Cancer Res Treat* 2017;164:119–31.
- [9] Boland GM, Chang GJ, Haynes AB, et al. Association between adherence to National Comprehensive Cancer Network treatment guidelines and improved survival in patients with colon cancer. *Cancer* 2013;119:1593–601.
- [10] Colorectal cancer overview – NICE Pathways. Available at: <https://pathways.nice.org.uk/pathways/colorectal-cancer> [accessed February 24, 2017].
- [11] Linee Guida – AIOM. Available at: <http://www.aiom.it/professionisti/documenti-scientifici/linee-guida/1,413,1> [accessed February 24, 2017].
- [12] Chung K-P, Chang Y-J, Lai M-S, et al. Is quality of colorectal cancer care good enough? Core measures development and its application for comparing hospitals in Taiwan. *BMC Health Serv Res* 2010;10:27.
- [13] Malin JL, Asch SM, Kerr EA, et al. Evaluating the quality of cancer care: development of cancer quality indicators for a global quality assessment tool. *Cancer* 2000;88:701–7.
- [14] Maria TG, Russo AG. Analisi dei percorsi diagnostico-terapeutici in oncologia: i tumori del colon e del retto. 2015. Available at: <http://www.epiprev.it/pubblicazione/epidemiol-prev-2015-39-4-allegato>. Accessed February 03, 2020.
- [15] Jacobs R, Goddard M, Smith PC. How robust are hospital ranks based on composite performance measures? *Med Care* 2005;43:1177–84.
- [16] Jha AK, Li Z, Orav EJ, et al. Care in US hospitals—the Hospital Quality Alliance program. *N Engl J Med* 2005;353:265–74.
- [17] Reeves D, Campbell SM, Adams J, et al. Combining multiple indicators of clinical quality: an evaluation of different analytic approaches. *Med Care* 2007;45:489–96.



- [18] AHRQ Quality Indicators: Composite Measures User Guide For The Patient Safety Indicators (PSI). 2010. Available at: [https://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V42/Composite\\_User\\_Technical\\_Specification\\_PSI.pdf](https://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V42/Composite_User_Technical_Specification_PSI.pdf).
- [19] Shwartz M, Ren J, Peköz EA, et al. Estimating a composite measure of hospital quality from the Hospital Compare database: differences when using a Bayesian hierarchical latent variable model versus denominator-based weights. *Med Care* 2008;46:778–85.
- [20] Shwartz M, Peköz EA, Christiansen CL, et al. Shrinkage estimators for a composite measure of quality conceptualized as a formative construct. *Health Serv Res* 2013;48:271–89.
- [21] Baker FB, Kim S-H. *Item Response Theory: Parameter Estimation Techniques*. CRC Press, 2004.
- [22] Van der Linden WJ, Hambleton RK. *Handbook of Modern Item Response Theory*. 2013; Springer Science & Business Media,
- [23] Chen T-T, Lai M-S, Lin I-C, et al. Exploring and comparing the characteristics of nonlatent and latent composite scores: implications for pay-for-performance incentive design. *Med Decis Making* 2012;32:132–44.
- [24] Normand S-LT, Wolf RE, McNeil BJ. Discriminating quality of hospital care in the United States. *Med Decis Making* 2008;28:308–22.
- [25] Teixeira-Pinto A, Normand S-LT. Statistical methodology for classifying units on the basis of multiple-related measures. *Stat Med* 2008;27:1329–50.
- [26] Bray F, Colombet M, Mery L, et al. *Cancer Incidence In Five Continents. Vol 11*. Lyon: International Agency for Research on Cancer; 2017.
- [27] Bland JM, Altman DG. Statistics notes: Cronbach's alpha. *BMJ* 1997;314:572.
- [28] Babbie ER. *The Basics of Social Research*. Cengage Learning, 2013.
- [29] Edwards JR, Bagozzi RP. On the nature and direction of relationships between constructs and measures. *Psychol Methods* 2000;5:155.
- [30] Thurstone LL. A method of scaling psychological and educational tests. *J Educ Psychol* 1925;16:433.
- [31] Andrich D. A rating formulation for ordered response categories. *Psychometrika* 1978;43:561–73.
- [32] Bartolucci F, Bacci S, Gnaldi M. *Statistical Analysis of Questionnaires: A Unified Approach Based on R and Stata*. CRC Press, 2015.
- [33] Cristobal E, Flavián C, Guinaliú M. Perceived e-service quality (PeSQ): measurement validation and effects on consumer satisfaction and web site loyalty. *Manag Serv Qual Int J* 2007;17:317–40.
- [34] Cerny BA, Kaiser HF. A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivar Behav Res* 1977;12:43–7.
- [35] Chien T-W, Chang Y, Wen K-S, et al. Using graphical representations to enhance the quality-of-care for colorectal cancer patients. *Eur J Cancer Care (Engl)* 2018;27:e12591.
- [36] Rasch G. *Probabilistic Models For Some Intelligence and Achievement Tests*. Copenhagen: Danish Institute for Educational Research; 1960.
- [37] *Care Patterns in Medicare and Their Implications for Pay for Performance*. NEJM. <https://www.nejm.org/doi/full/10.1056/nejmsa063979> [accessed July 24, 2018].