Global Leadership Initiative on Malnutrition (GLIM)

GLIM Consortium; Keller, Heather; de van der Schueren, Marian A. E.; Jensen, Gordon L.; Barazzoni, Rocco; Compher, Charlene; Correia, M. Isabel T. D.; Gonzalez, M. Cristina; Jager-Wittenaar, Harriet; Pirlich, Matthias

Published in:
Journal of Parenteral and Enteral Nutrition

DOI:
10.1002/jpen.1806

Review

Journal of Parenteral and Enteral Nutrition
Volume 44 Number 6
August 2020 992–1003

DOI: 10.1002/jpen.1806
wileyonlinelibrary.com
WILEY

Global Leadership Initiative on Malnutrition (GLIM): Guidance on Validation of the Operational Criteria for the Diagnosis of Protein-Energy Malnutrition in Adults

Heather Keller, RD, PhD[1,2]; Marian A. E. de van der Schueren, RD, PhD[3], for the GLIM Consortium; Gordon L. Jensen, MD, PhD[4]; Rocco Barazzoni, MD, PhD[5]; Charlene Compher, PhD, RD[6]; M. Isabel T. D. Correia, MD, PhD[7]; M. Cristina Gonzalez, MD, PhD[8,9]; Harriët Jager-Wittenaar, PhD, RD[10,11]; Matthias Pirlich, MD, PhD[12]; Alison Steiber, PhD, RDN[13]; Dan Waitzberg, MD, PhD[14]; and Tommy Cederholm, MD, PhD[15,16]

Abstract

Background: The Global Leadership Initiative on Malnutrition (GLIM) created a consensus-based framework consisting of phenotypic and etiologic criteria to record the occurrence of malnutrition in adults. This is a minimum set of practicable indicators for use in characterizing a patient/client as malnourished, considering the global variations in screening and nutrition assessment, and to be used across different healthcare settings. As with other consensus-based frameworks for diagnosing disease states, these operational criteria require validation and reliability testing, as they are currently based solely on expert opinion. *Methods:* Several forms of validation and reliability are reviewed in the context of GLIM, providing guidance on how to conduct retrospective and prospective studies for criterion and construct validity. *Results:* There are some aspects of GLIM that require refinement; research using large databases can be employed to reach this goal. Machine learning is also introduced as a potential method to support identification of the best cut points and combinations of indicators for use with the different forms of malnutrition, which the GLIM criteria were created to denote. It is noted as well that validation and reliability testing need to occur in a variety of sectors and populations and with diverse persons using GLIM criteria. *Conclusion:* The guidance presented supports the conduct and publication of quality validation and reliability studies for GLIM. (*JPEN J Parenter Enteral Nutr.* 2020;44:992–1003)

Keywords

adult; nutrition assessment; outcomes research/quality

From the [1]Schlegel-University of Waterloo Research Institute for Aging, Waterloo, Ontario, Canada; [2]Department of Kinesiology, University of Waterloo, Waterloo, Ontario, Canada; [3]Department of Nutrition and Health, School of Allied Health, HAN University of Applied Sciences, Nijmegen, the Netherlands; [4]The Larner College of Medicine, University of Vermont, Burlington, Vermont, USA; [5]Department of Medical, Surgical and Health Sciences, University of Trieste, Trieste, Italy; [6]Biobehavioral Health Sciences Department, School of Nursing, University of Pennsylvania, Philadelphia, Pennsylvania, USA; [7]Medical School, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil; [8]Post-Graduate Program in Health and Behaviour, Catholic University of Pelotas, Pelotas, Rio Grande do Sul, Brazil; [9]Post-Graduate Program in Nutrition and Food, Federal University of Pelotas, Pelotas, Rio Grande do Sul, Brazil; [10]Research Group Healthy Ageing, Allied Health Care and Nursing, Hanze University of Applied Sciences, Groningen, the Netherlands; [11]Department of Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands; [12]Endocrinology, Gastroenterology, Clinical Nutrition, Imperial Oak Outpatient Clinic (Kaisereiche), Berlin, Germany; [13]Academy of Nutrition and Dietetics, Chicago, Illinois, USA; [14]Department of Gastroenterology, School of Medicine, University of Sao Paulo, Sao Paulo, Brazil; [15]Clinical Nutrition and Metabolism, Department of Public Health and Caring Sciences, Uppsala University, Uppsala, Sweden; [16]Theme Aging, Karolinska University Hospital, Stockholm, Sweden.

Financial disclosure: None declared.

Conflicts of interest: Authors are GLIM working group members.

This article is simultaneously published by The European Society for Clinical Nutrition and Metabolism in the journal Clinical Nutrition and by the American Society for Parenteral and Enteral Nutrition in the Journal of Parenteral and Enteral Nutrition and will be subsequently published by The Society on Sarcopenia, Cachexia and Wasting Disorders in the Journal of Cachexia, Sarcopenia and Muscle. Minor differences in style may appear in each publication, but the article is substantially the same in each journal.

Received for publication September 23, 2019; accepted for publication January 22, 2020.

This article originally appeared online on June 12, 2020.

Corresponding Author:
Heather Keller, RD, PhD, Schlegel-University of Waterloo Research Institute for Aging, Waterloo, Ontario, Canada.
Email: hkeller@uwaterloo.ca

Clinical Relevancy Statement

Malnutrition across care sectors throughout the world impacts patient outcomes and healthcare utilization. Prevalence rates vary because of the use of different tools. To promote global communication and indexing of malnutrition in the International Statistical Classification of Diseases, a common global framework is needed to categorize patients as malnourished. The Global Leadership Initiative on Malnutrition (GLIM) operational criteria provide such a framework. As the GLIM criteria are based on consensus, further evidence is required to validate and test GLIM for reliability. Best practices and guidance for this testing are provided. Criterion validation specifically is needed and feasible within the global context. The use of large datasets for this criterion validation will support the refinement of the GLIM operational criteria.

Background

Clinical nutrition researchers for more than a decade have been interested in developing common terminology and criteria for the identification of malnutrition across global healthcare sectors. Over time, the malnutrition definition has changed because of a greater knowledge and awareness of the role of inflammation and the impact that body composition derangements play on outcomes related to this condition. Malnutrition is defined here as a subacute or chronic state of nutrition, in which a combination of varying degrees of undernutrition or overnutrition and inflammatory activity have led to changes in body composition and diminished function.[1-4]

Malnutrition may be corrected or improved by interventions that enhance food/nutrient intake and meet requirements,[5] but in the presence of inflammation, such as in disease-related malnutrition, the benefits of nutrition treatment may be blunted.[6] International guideline committees have agreed on the etiologic basis of the undernutrition form of malnutrition as resulting from decreased intake and/or assimilation of energy and protein intake and/or inflammation resulting in catabolism of lean tissues, with further categorization of malnutrition associated with (1) chronic disease or conditions with sustained inflammation, (2) chronic disease with minimal or no perceived inflammation, (3) acute disease or injury with severe inflammation, or (4) pure chronic starvation not related to disease.[3,7] The Global Leadership Initiative on Malnutrition (GLIM) used a consensus-based approach to produce operational criteria for these various forms of malnutrition using a minimum of 1 phenotypic and 1 etiologic component. These are known as the "GLIM criteria."

The GLIM operational diagnostic criteria are based on the minimum phenotypic and etiologic indicators of significant weight loss or low body mass index (BMI) or low muscle mass and reduced food intake or its assimilation or inflammation.[8,9] These are operational criteria in that they operationalize the consensus-based definitions of malnutrition and the more specific subcategories based on the etiology of impaired nutrition intake, reduced assimilation, and exacerbation of these by inflammation due to chronic or acute disease. Significant weight loss, low BMI, and/or low skeletal muscle mass are clinical features expressing (the severity of) malnutrition.

It is important to provide some points of clarification on the GLIM criteria. First, these criteria were designed to be used globally, in different healthcare sectors, to provide a common basis for diagnosis; this is not a "new" definition of malnutrition. However, it is unclear whether the criteria are valid in all of these scenarios. Further, they are a "minimum criteria" for the diagnosis of malnutrition globally, recognizing the varying capacities and traditions of regions and healthcare systems for the conduct of nutrition assessment to diagnose malnutrition. Also, these operational criteria do not identify micronutrient malnutrition but only protein-energy malnutrition and, specifically, undernutrition. Finally, they are based on consensus of experts, who—although world leaders—have the potential for bias that influenced the development of these criteria. As with any new tool, validation of the GLIM criteria is important to support their dissemination and uptake into practice.

This paper will provide a brief overview of how the GLIM criteria were developed, guidance on how they should be positioned for use in nutrition care, and how they are to be validated and tested for reliability with retrospective or prospective studies.

Overview of GLIM Criteria and Development

The GLIM operational diagnostic criteria were developed over a period of 3 years (2016–2018) by collective leadership of 4 major clinical nutrition societies (American Society for Parenteral and Enteral Nutrition [ASPEN], European Society for Clinical Nutrition and Metabolism [ESPEN], Federación Latinoamericana de Terapia Nutricional, Nutrición Clínica y Metabolismo [FELANPE], Parenteral and Enteral Nutrition Society of Asia [PENSA]). Representatives from diverse disciplines, considered leaders in their societies, were invited to participate and act for their societies. At face-to-face meetings, these leaders shared their region's current practices and challenges in the diagnosis of malnutrition; developed a common understanding of the phenotype and etiology of malnutrition using extant research; compared and contrasted screening and assessment tools for common malnutrition indicators; and identified and ranked by secret ballot the minimum indicators that should be included.[8,9] As such, GLIM criteria are consensus-based, and validation is needed to confirm and, if necessary, refine these

operational criteria. Within the medical field, this is an accepted process; evidence-informed consensus has been used to develop diagnostic criteria, which are then validated and refined at regular time points (eg, Alzheimer's disease dementia,[10] multiple sclerosis,[11] and sarcopenia[12]).

Cut points for defining a significant weight loss, low BMI, and reduced food intake were based on prior evidence when available (eg, weight loss, BMI) or based on the best judgment of the GLIM group. Details on how to categorize inflammation, low muscle mass, and reduced assimilation of food intake due to gastrointestinal conditions are not yet provided because of a lack of clearly defined cut points or markers. Phenotypic indicators are used to determine severity of malnutrition. Validation studies during the next few years will be crucial in determining whether the suggested framework works in practice and will help to develop cut points and further definition for disease state/inflammation, muscle mass assessment, and assimilation of food and in evaluating whether malnutrition severity indicators are appropriate.

Positioning GLIM Criteria Within Nutrition Care

As outlined in the GLIM consensus report, screening and assessment processes occur outside of, and feed into, the GLIM operational diagnostic criteria.[8,9] GLIM does not replace current validated screening and assessment tools but rather is used alongside these tools to offer minimum criteria for the classification or description of a patient as malnourished. As noted above, GLIM is neither comparable to nor replaces a comprehensive nutrition assessment in clinical care. These minimum operational criteria are needed globally to speak one language across the world, to understand the variation in prevalence of protein-energy malnutrition among regions and populations, and to support the development of an updated International Classification of Diseases coding for malnutrition.[8,9]

The consensus-based definition of risk screening is "a rapid process performed to identify subjects at nutritional risk."[3] There is a wide variety of valid and reliable screening tools often created for specific populations or healthcare settings. These tools may identify patients already malnourished or at risk of malnutrition or assess risk factors that may result in malnutrition in the future. As a rapid tool, they typically require minimal expertise and only include a few nutrition indicators or risk factors. Assessment, on the other hand, provides a "basis for the diagnosis decision, as well as for further actions including nutritional treatment."[3] A clinical nutrition assessment completed by a trained healthcare professional (eg, dietitian, physician) has many components, including adequacy of food intake in comparison with nutrition requirements, functional ability (and growth for children), clinical/medical history, physical exam, weight history, body composition, and biochemical assessment,[4,13] with specific components, standards, and references often particular to a population group (eg, head circumference in children). A comprehensive assessment can lead to a diagnosis of undernutrition, overnutrition, or micronutrient deficiency or excess. Because of its in-depth nature, a comprehensive nutrition assessment provides details that are needed to understand the root causes of the malnutrition, guide interventions and monitoring, and anticipate outcomes based on severity of malnutrition. To standardize clinical practice and advance research in malnutrition, a few valid and reliable assessment tools have been created that are more comprehensive than screening tools but less thorough than a complete clinical nutrition assessment (eg, Subjective Global Assessment [SGA],[14] Mini Nutritional Assessment [MNA] [for older adults],[15] and Patient-Generated SGA [PG-SGA] [for patients with cancer[16,17]]). These tools specifically target protein-energy malnutrition, as this has been the greatest concern with respect to healthcare use in much of the developed world.

When developing GLIM, emphasis was placed on indicators that would reflect the protein-energy malnutrition definition[1,3] while being appropriate for diverse settings and contexts (eg, outpatient clinics, intensive care unit, acute care, residential care, community) and throughout the world (in both high-income and low-income countries, applicable to different ethnicities). To fulfill the ideal that the GLIM criteria should be applicable everywhere, without the need for (expensive) diagnostic equipment, practicability of included parameters was a prerequisite.[18,19] Because of the dichotomous nature of several of the indicators used in the GLIM criteria and timeframes (eg, 6 months' weight change), they are not anticipated to be useful to monitor changes in nutrition status posttreatment, nor was this the intent of GLIM.

Guidance on Validation of GLIM Criteria

As noted in the consensus reports,[8,9] the next step in this initiative is to validate and further detail GLIM. Validity is the extent to which GLIM criteria identifies what it is intended to identify, that is, protein-energy malnutrition. (Note: the term "identify" is used here rather than "measure," as GLIM is not explicitly a measurement tool but a diagnostic framework based on several measures.) It should be noted that this is quite different from a tool that predicts outcomes known to be associated with malnutrition (eg, length of stay). Table 1 provides an overview of types of validation and a hierarchy that can be used to test validity.

There are several forms of validation,[20-22] with the lowest being face or content validity and the best form of validation being concurrent criterion validity, which is determined by comparing a test tool to a "gold standard." GLIM criteria are considered to have face validity in that they

Table 1. Definitions of Types of Validity and Guidance on When and How Various Forms Can Be Assessed.

Type of Validity	Definition	Detailed Description	Methodological Considerations
1. Criterion	Measures or identifies what it is intended to.	When a measure is considered by the field to represent or diagnose the health condition, it is described as the gold standard. Comparison of the test measure to this gold standard demonstrates criterion validity and is the ultimate form of validity, as it confirms that the new tool identifies what it is intended to. In the case of GLIM, the gold standard would need to identify protein-energy malnutrition, as this is what GLIM was designed to identify.	See Concurrent and Predictive Validity.
1a. Concurrent	The test measure is compared with a gold-standard measure (ie, criterion) that is collected at the same point in time.	This is one form of criterion validity that can only be established with an accepted gold standard for malnutrition.	The test measure is compared with the gold standard, which is collected concurrently. It can be completed in relatively few studies, based on different populations.
1b. Predictive	Ability of the test measure to predict a future outcome.	This is another form of criterion validity, but not as confirmatory as a gold-standard criterion comparison. When a gold-standard criterion is not available, this form of validity is a substitute but needs construct validity to further demonstrate the relevance of the tool. In the case of GLIM, this would be in comparison to meaningful health outcomes that are expected to be associated with protein-energy malnutrition.	Predictive validity is often done in a variety of studies to demonstrate the significance of the test measure when compared with health outcomes. It is important to choose health outcomes relevant to the population and sector, as well as the construct of interest; in this case of GLIM, the construct of interest is protein-energy malnutrition.
2. Construct	The test measure is associated with other health constructs or similar tools in the way that is anticipated.	When there is no gold standard, construct validity is necessary to demonstrate the value of the test tool for measuring the construct of interest (ie, in the case of GLIM, protein-energy malnutrition). If a gold standard exists, construct validity is less of a priority and considered inferior for determining the value of the test tool.	Hypotheses are formulated on how GLIM should be highly correlated with relevant measures/instruments and not with irrelevant measures. A form of construct validity (discriminant) would be if GLIM prevalence differs among groups in which prevalence is anticipated to vary (eg, hospital vs community). Convergent validity, for example, would be finding a positive association between health-related quality of life and GLIM criteria. Convergent validity is typically based on other health measures collected at the same time as GLIM criteria.

(*continued*)

Table 1. (continued)

Type of Validity	Definition	Detailed Description	Methodological Considerations
3. Content and face	The test measure includes relevant concepts and indicators. This is the lowest form of validity.	This form of validity is established first. Face validity is often completed by experts or knowledgeable practitioners who ensure wording and content are consistent with the concept being assessed. In the case of GLIM, criteria are relevant to protein-energy malnutrition. Content validity is observed when a new tool "covers" all of the subcomponents of the construct. Face validity confirms that the tool includes measures that are consistent with the construct.	This form of validity is often based on the development process for the new tool. An evidence base and experts are used to identify relevant components and consensus among experts used to derive key components.

GLIM, Global Leadership Initiative on Malnutrition.

were developed by an evidence-informed consensus group of experts, including voting on which nutrition indicators should be included.[8,9]

Although the experts were appointed by the 4 nutrition societies involved and were selected to represent different professions, disciplines, and regions, their appointments may have had the potential to introduce bias, as they were all selected and agreed to work toward developing a universal framework. This is why face and content validity are insufficient for a diagnostic framework like GLIM. Face validation with typical practitioners was not formally completed, although society representatives may have vetted early versions with their peers. Indicators used are considered relevant to the diagnosis of malnutrition and, for the most part, are acceptable and practicable to end users.[21] Content validity is focused on the inclusion of all relevant concepts and exclusion of irrelevant concepts as judged by experts[19]; this is done by including key questions or items that represent subcomponents of the concept. To confirm validity, these selected items need to demonstrate that they represent the domain of interest (eg, inclusion of inflammatory conditions adequately represents inflammation in the GLIM criteria).[19] In the case of GLIM, key components crossing diverse areas of nutrition assessment (eg, weight, inflammation, food intake) are included, but confirmation of their value in representing the subcomponent is needed.

The preferred form of validation is criterion validity, comprising both concurrent and predictive validity. In this situation, there is an understanding of what the concept of malnutrition is, and it can be measured in a valid and reliable way. A gold standard is sometimes the term used to describe a criterion. In the absence of a worldwide gold standard for malnutrition, an in-depth nutrition assessment completed by a trained nutrition expert is regarded as a semi-gold standard[23] and the preferred tool for criterion validation. Nutrition assessment comprises many aspects related to nutrition status, including assessment of dietary intake in relation to requirements; weight history; biochemistry; body composition; factors impeding nutrition intake; physical, psychological, and social well-being; disease history; and financial status.[24,25] SGA or MNA (for older adults), which are brief, standardized tools, have been validated against a nutrition assessment by a clinician[14,15]; they are thus considered "fuzzy, semi-gold" standards and will be appropriate but less conclusive at determining validity of GLIM than a comprehensive nutrition assessment.[23,26] Concurrent criterion validity would be the collection of GLIM indicators at the same time as the completion of the (semi–)gold-standard criterion. As GLIM is essentially a minimum list of key indicators or variables to identify malnutrition, there may be special challenges in criterion validation. Specifically, there is a high likelihood for criterion contamination if the GLIM indicators are already embedded into the criterion measure being used. Prior work has criticized a variety of screening tools for this issue.[18,23]

Often, a gold or semi-gold standard is not available, and prediction of a meaningful health outcome that is known to be associated with malnutrition can be used; this is called predictive criterion validity.[19] In this form of criterion validity, GLIM would be expected to predict a meaningful outcome that is known to be associated with protein-energy malnutrition (eg, 6-month mortality). It is important to note, however, that predictive validity for meaningful health outcomes does not confirm that GLIM identifies the

Table 2. Potential Meaningful Health Outcomes to Be Used as the Comparator in Validation Studies.

Healthcare Setting	Health Outcome
Hospital	In-hospital mortality
	Major complications
	30-day mortality
	30-day readmission rate
	60-day readmission rate
	Length of hospital stay
Nursing home	3-month mortality
	1-year mortality
	Quality of life
	Functionality
Community	Admission rate to hospital or nursing home
	Healthcare use (eg, Physician visits, hospital)
	Functionality
	Quality of life

construct of interest (ie, protein-energy malnutrition) but that it is identifying something that is associated with the health outcome chosen in the analysis. Health outcomes for use in predictive validity may depend on the setting in which GLIM is used. For example, in the hospital sector, 30- or 60-day readmission rates are relevant outcomes, whereas in the community sector, it may be 12- or 36-month hospital admission or mortality rates. Table 2 provides a list of health outcomes for different sectors that may be most relevant.

Construct validity is completed when there is no gold standard or criterion for comparison and a longitudinal study for prediction of an outcomes is not possible; this is considered an indirect form of validation.[21] For example, construct validity can involve comparing GLIM criteria to measures of frailty. It would be anticipated that if GLIM criteria were identifying something important that is believed to be protein-energy malnutrition, it should be associated with valid and reliable measures of frailty, as malnutrition and frailty are often comorbid. This is an example of convergent validity (hypothesized to be associated with the other measure in the expected direction). Divergent validity (hypothesized to not be associated with the other measure) is not as commonly determined. An example would be finding no association between GLIM and body height; it is anticipated in most healthcare sectors in the developed world that the height of an adult patient is not associated with acute malnutrition. Discriminant validity means GLIM can discriminate among groups in which a gradient in the prevalence of malnutrition is expected. An example of discriminant validity would be having a high prevalence of malnutrition using the GLIM criteria in hospital patients and a lower prevalence in community samples, as it is well known that the prevalence is higher in acute care. These are all forms of construct validity.[21] Validity is not absolute but is a "matter of degree."[21] Ongoing assessment of validity and several forms of validity are recommended to confirm the utility of GLIM.[22]

Reliability of GLIM Criteria

Reliability is the degree to which the results obtained by a measurement or procedure can be replicated, either by the same assessor or different assessors.[27] Reliability of individual nutrition indicators that make up the GLIM criteria is generally available (eg, weight, BMI, presence or absence of a disease condition) and adequate, if established protocols for data collection and manipulation (eg, calculation of percent weight change) are followed. Laboratory measures of inflammation should be considered when rigorous validation testing is the objective to promote reliability; however, underlying medical diagnosis may guide inflammatory assessment in clinical settings. For some GLIM indicators, reliability is likely to be dependent on the measurement used. For example, determining muscle mass with anthropometry is less reliable than dual-energy absorptiometry because of interobserver error, especially for those untrained in performing anthropometry measurements.[28] The reliability of food-intake assessment is dependent on how this intake is determined. As GLIM requires an understanding of recent food intake, a food diary or a 24-hour recall (reflecting recent intake) is more appropriate than a food frequency questionnaire (reflecting intake over, eg, the past 6 months).[13] At minimum, when conducting a validation study on GLIM, authors should report how indicators were measured and the reliability of indicator determination.

In addition to the underlying reliability of the indicators used in GLIM, because of the nature of GLIM scoring, interrater reliability on categorization of malnutrition using GLIM criteria should also be determined. When reviewing the same patient data, would 2 clinicians/researchers/data extractors identify the same phenotypic and etiologic indicators as occurring or being triggered (yes/no)? To support this reliability, greater detail on what body composition measures should be used and the specific cut points required for determining low muscle mass are needed. There is guidance in GLIM on the gastrointestinal conditions and symptoms that can lead to reduced assimilation of nutrients, as well as inflammatory diagnoses to consider as acute or chronic. However, the GLIM criteria require clinical judgment to determine severity of these conditions and whether the indicator is triggered. Guidance on intensity, frequency, and duration of symptoms is needed to support interrater reliability.

Furthermore, validity and reliability need to be established in diverse patient groups and sectors for GLIM to be used globally[22,27]; for example, it cannot be assumed that

if GLIM is found to have criterion validity in hospital patients, it would also be valid for community-based patients. The latter needs to be demonstrated in its own validation study.

Study Design and Considerations in Validating GLIM and Testing Interrater Reliability

An initial consideration in designing a validation study for GLIM is the determination of a criterion or constructs appropriate for comparison. Constructs should theoretically be related to the concept of malnutrition (eg, quality of life). As previously noted, a gold standard for malnutrition is elusive, but the MNA Full Form (MNA-FF), SGA, and PG-SGA for oncology patients and/or comprehensive nutrition assessment by a trained clinician have been identified to be relevant semi-gold criteria for validation studies.[23,29] Individual indicators of nutrition status such as biochemical parameters, anthropometry, or food intake are insufficient for criterion validation.[18,23,29] Similarly, combined scores of various tools assembled specifically for a validation study (and thus having unclear validity themselves) are inappropriate for GLIM validation studies.[18,23] Related to the identification of an appropriate criterion is the requirement that the GLIM criteria being tested are not embedded within the criterion being used to assess GLIM. This is considered criterion contamination or incorporation bias.[18,20,22,23] Assessors of nutrition status should be blinded to diagnostic GLIM criteria results. Finally, when a criterion has unknown reliability, 2 assessors, blinded to the results of the GLIM diagnostic categorization, should evaluate participants on their nutrition status and be blinded to each other's assessment.[22]

Once the criterion or construct for comparison has been selected, the next step is to determine sample size. For retrospective studies, determining whether the available sample provides sufficient power to estimate statistics with sufficient precision can be determined with a back-calculation. Jones[22] provides guidance on sample-size calculation for both construct and criterion-validation studies. Participants in either a retrospective or prospective validation study should be included irrespective of their nutrition status so as to not bias the results.[22]

Since validation of the GLIM criteria can be tested by both retrospective and prospective studies, each study design has specific aspects to take into consideration. It is recommended that retrospective validation studies be based on at least 1 phenotypic and 1 etiologic GLIM indicator, and for prospective studies, all GLIM variables be included. Such a strategy will support comparison on the prevalence of malnutrition using various combinations of the etiologic and phenotypic indicators included in GLIM as well as comparison of validity statistics.[19] Validation studies that are based on phenotypic indicators but missing etiologic variables, or the other way around, are not considered sound validation studies, as they violate the underlying concept of GLIM that malnutrition diagnosis should be made on their combination. Also, use of a medical diagnosis that may impact assimilation of nutrients or cause inflammation without a severity rating (ie, assuming that all patients with, for example, cancer will have inflammation, or all patients with Crohn's disease will have reduced assimilation) is insufficient for validation testing. In prospective studies, severity of disease needs to be measured in a reliable and valid manner. Furthermore, the various indicators for severity of diagnosis and cut points for muscle mass and inflammatory markers can be tested to determine which is best for use as GLIM criteria, with respect to sensitivity (SE) and specificity (SP). Machine learning (eg, random forest) in large datasets could be useful for identifying the many different cut points and GLIM variables that could be used in combination.[30] Machine learning uses algorithms and statistical models when a wealth of high-quality data are available. This method could be used to validate cutoff points.

A minimum dataset for retrospective studies would be healthcare setting; country and continent; demographics (sex, age, ethnicity); health characteristics; at least 1 phenotypic and 1 etiologic GLIM indicator with defined cut points; valid and reliable measurement of a construct that is theoretically associated with malnutrition or inclusion of a concurrent semi–gold-standard criterion or predictive health outcome (not concurrent with determination of GLIM). When the gold-standard criterion has some subjective component (eg, nutrition assessment) and has no determination of reliability (eg, not assessed by >1 assessor for the test patient), this should be noted as a limitation of the validation. A sufficient sample size would be demonstrated using the appropriate methods for estimation for construct or criterion validity.[22] Ideally, reliability of variables used in GLIM and the construct for comparison are determined in the validation sample.

The minimum dataset for prospective criterion-validation studies includes healthcare setting; country and continent; demographics (sex, age, ethnicity); health characteristics; determination of measures used in GLIM; inclusion of valid and reliable severity rating for assimilation of food and inflammatory diseases; a variety of muscle mass measures for comparison, especially for those considered less reliable and accurate (eg, bioelctrical impedance analysis, anthropometry); an in-depth assessment of nutrition status (or, less preferably, SGA or PG-SGA in oncology patients or MNA-FF), preferably completed by 2 assessors independently; and health outcomes that are relevant for prediction. Reliability testing of some GLIM indicators that lack precision (eg, food intake) should also be considered. When these design components are missing, this should be noted in the limitations of the

Table 3. Recommendations for Validation Studies of GLIM Criteria.

Type of Validation Study	GLIM Criteria	Validity Comparator	Minimum Other Data Required
Retrospective validation studies: Criterion or construct validity	Measured data including at least 1 phenotypic and at least 1 etiologic indicator.	Semi-gold standard for malnutrition (preferably in-depth nutrition assessment by a trained professional, MNA-FF, SGA/PG-SGA: or malnutrition-related health outcome, such as in-hospital complication or mortality) or a valid measurement of a construct that is known to be related to malnutrition.	Healthcare setting; country and continent; demographics (sex, age, ethnicity); health characteristics.
Prospective validation studies: Criterion validity	Measured data of all phenotypic and etiologic indicators, preferably including a variety of muscle mass measurements for reasons of comparison, a clear description of measurements of inflammation, and food-intake/assimilation assessment.	Semi-gold standard for malnutrition, preferably in-depth nutrition assessment by a professional trained in nutrition, performed independently by 2 trained assessors; or MNA-FF, SGA/PG-SGA; and malnutrition-related health outcome, such as in-hospital complication or mortality.	Healthcare setting; country and continent; demographics (sex, age, education level, ethnicity); health characteristics.

GLIM, Global Leadership Initiative on Malnutrition; MNA-FF, Mini Nutritional Assessment Full Form; PG-SGA, Patient-Generated SGA; SGA, Subjective Global Assessment.

study. Prospective studies provide an ideal opportunity to test a variety of cut points for GLIM variables. The recommendations for validation studies of GLIM are summarized in Table 3.

Statistical Analyses for Reliability and Validity of GLIM Criteria

Interrater reliability for triggering GLIM indicators is recommended to support use of these operational criteria, as some components are subjective (eg, severity of disease that causes inflammation). Each variable or indicator can be present or absent in the scoring of GLIM. As this is a dichotomous response, κ would be used to determine interrater reliability and agreement among raters. κ accounts for the chance event that agreement is observed.[22] κ could be calculated for each indicator as well as for the overall categorization of "malnourished," using combinations of 1 phenotypic and 1 etiologic variable included in the study. κ that is >0.80 is substantial, whereas 0.61–0.80 is moderate[31]; any lower values bring into question the reliability of the GLIM criteria. The 95% confidence interval is as important and more informative than the κ value and should also be presented.[22,31]

When various GLIM variables are considered in a single validation study, descriptive analyses demonstrating prevalence of malnutrition based on specific combinations of phenotypic and etiologic indicators should be completed. This could include not only the recommended 1 phenotypic and 1 etiologic indicator but also other combinations, for example, 2 phenotypic and 1 etiologic, or the same phenotypic indicator with another etiologic indicator, to test and refine GLIM. Again, machine learning would help to identify all possible combinations in a dataset for comparison to the gold standard (Figure 1[32]). Prevalence for these combinations of GLIM contrasted with the prevalence as determined by the gold standard used in the validation study would also be an important result to document.

Construct validity is based on the convergent association of GLIM with constructs that it should be associated with, if GLIM sufficiently identifies malnutrition. Comparative analyses are typically done to determine construct validity. For example, discriminant validity would be determined by comparing prevalence of malnutrition based on GLIM criteria in various healthcare sectors. Proportions that are identified with malnutrition could be contrasted using a *z*-test. Convergent validity is typically determined by association analyses. Is the GLIM categorization of malnutrition associated with health-related outcomes, such as physical

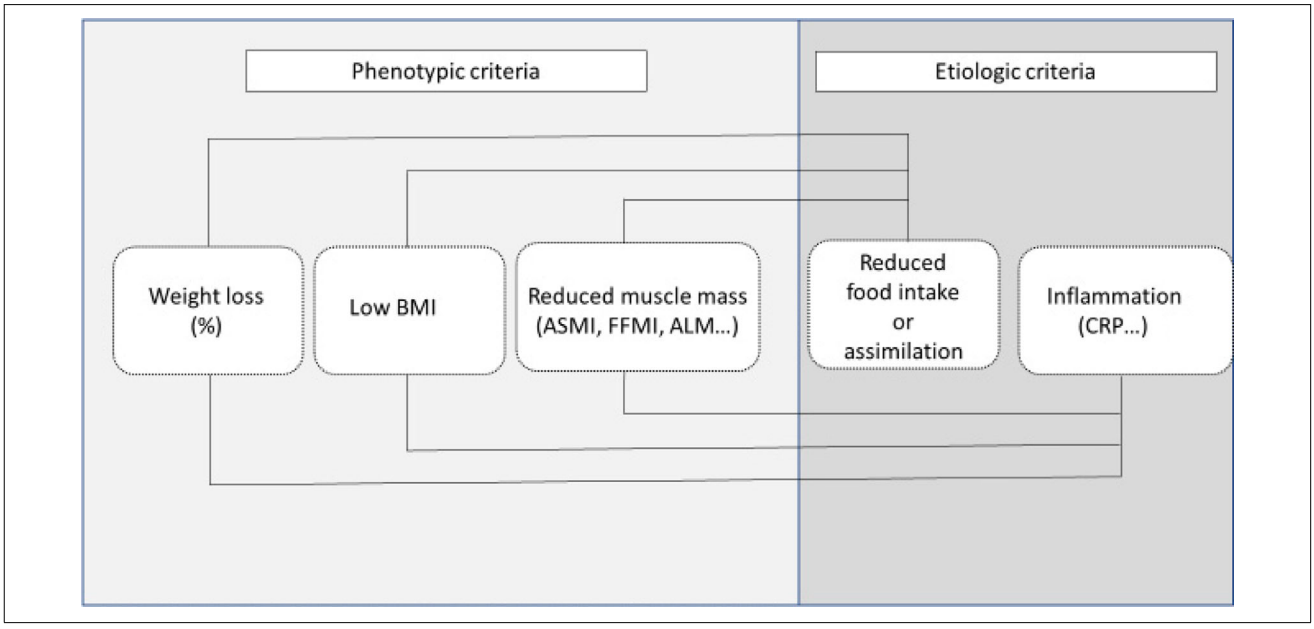

Figure 1. Potential combinations of GLIM criteria for validation. ALM, appendicular lean mass; ASMI, appendicular skeletal muscle index; BMI, body mass index; CRP, C-reactive protein; FFMI, fat-free mass index; GLIM, Global Leadership Initiative on Malnutrition. Figure adapted with permission from Reference 32.

functioning or quality of life? These types of analyses have been done with the ESPEN definition of malnutrition.[33] Bivariate tests will determine this association (eg, χ^2, analysis of variance). It is important to note that construct validity is only established with respect to the constructs with which GLIM is compared in these analyses.[22] Associations with various constructs (bivariate or multivariate) are insufficient for determining criterion validity. Associations simply indicate that 1 concept/measure coincides with another; it does not confirm the validity of the test measure for measuring a specific concept. With respect to predictive validity, when a health outcome is being predicted, association analyses are commonly used. The variance (ie, R^2) explained by GLIM would provide information on the relative importance of the GLIM categorization of malnutrition to the outcome, when adjusting for covariates. However, predictive multivariate analyses are specific for the sample on which they are conducted, and >1 study confirming GLIM to predict health outcomes with that population group (eg, hospital patients) is needed. Covariates should include demographics, disease severity, and other known predictors of the health outcome variable. Odds ratios or hazard ratios can be applied to determine predictive validity in relation to outcome parameters such as mortality, length of stay, or readmission.

The preferred statistical test for criterion validity is determination of SE and SP, as well as positive predictive values (PPVs) and negative predictive values (NPVs) (see Figure 2). Receiver operating characteristic curves are not applicable,

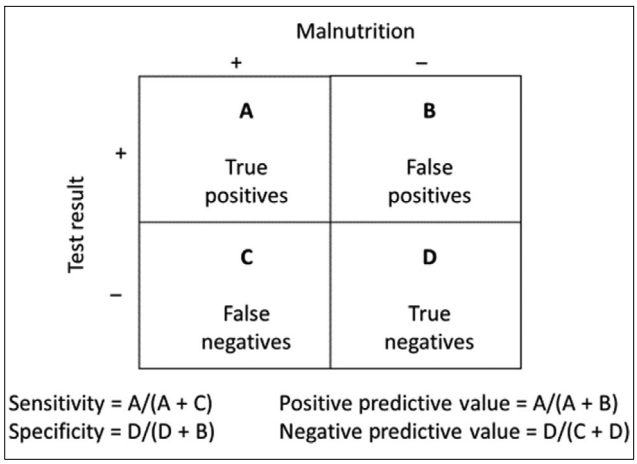

Figure 2. Calculating sensitivity, specificity, positive predictive value, negative predictive value.

as GLIM indicators are dichotomous (yes/no); however, such analyses would be beneficial to confirm the cut points used for GLIM variables (eg, what percent of weight loss is predictive of negative outcomes). If an outcome or criterion is numerical (eg, length of stay, MNA score), a cut point to indicate worse/better status on the outcome will be needed to determine SE and SP. Published cut points or, alternatively, the median could be used to dichotomize the outcome variable. Simple equations (Figure 2) are used to determine SE and SP, which can also be determined from the statistical output of a χ^2 test.

Table 4. Recommended Validation Statistics.

Test Statistic	Type of Validity/Type of Variable	Recommended Interpretation
Sensitivity	Criterion—categorical variables	>80% required
Specificity	Criterion—categorical variables	>80% required
χ^2	Construct—categorical variables	$P < .05$ if sample size is <200, $P < .01$ if sample size is ≥200 and presents 95% CI; large samples also use Cramers V (Φ_c)
t-Test	Construct—GLIM categorical variable and construct numerical	$P < .05$ if sample size is <200, $P < .01$ if sample size is ≥200 and presents 95% CI; samples also use Cohen's d (*d*)
Odds/hazard ratio	Predictive validity—categorical health outcome variable	≥2.0 required
z-Test	Construct validity—discriminant, proportion identified as malnourished per GLIM	$P < .05$ if sample size is <200, $P < .01$ if sample size is ≥200 and 95% confidence intervals
	Reliability	
κ	Agreement	>0.8 required

GLIM, Global Leadership Initiative on Malnutrition.

In addition to SE and SP, PPV and NPV provide insight into the utility of the test measure. PPV and NPV are influenced by the true prevalence of the condition[34] and thus will vary depending on reference method, the study population, and healthcare sector where GLIM is used. PPV and NPV provide insight into the utility of a tool considering this prevalence.

Although *P*-values also give an indication of statistical significance of a finding (for example, when studying malnutrition in relation to mortality risk), *P*-values give no indication of the effect size, because *P*-values are also influenced by the sample size. Thus, we advise reporting coefficients, 95% confidence intervals, and *P*-values, as well as use of Cohen's d and Cramer's V for large samples.[35] Table 4 outlines accepted values for association and SE and SP when conducting construct and criterion validity, respectively. Other common statistics recommended for outcomes and predictive validity are also provided.

Challenges for Refinement of GLIM Through Future Validation Studies

The GLIM framework is not considered static. The GLIM criteria are unique in that they provide, for the first time, a worldwide consensus for categorizing malnutrition. It is, however, recognized that modifications may be expected with updated versions in the future. The original GLIM paper identifies some points that need to be worked on during the next few years. At least 21 possible combinations of parameters (at least 1 phenotypic and 1 etiologic indicator) can be used to categorize a patient as malnourished or not (Figure 1), and it is of utmost interest to study which combination of variables (and with which cut points) GLIM meets its different goals of identifying those who are malnourished, predicting outcomes associated with a poor nutrition status, and/or predicting those who will respond to nutrition interventions. Also, it may be important to identify which of the GLIM variables contributes most to malnutrition prevalence in different subgroups of patients. Based on the taxonomy of malnutrition features—that is, (1) chronic disease or conditions with sustained inflammation, (2) chronic disease with minimal or no perceived inflammation, and (3) acute disease or injury with severe inflammation, or pure chronic starvation not related to disease[3,5]—it may well be that different combinations of etiologic and phenotypic indicators should be used for different forms of malnutrition or to achieve different goals. Responsiveness to a nutrition intervention may, for example, be different for patients who present with loss of body weight in combination with poor nutrition intake (reflecting starvation) compared with patients who also show low muscle mass and inflammation (reflecting disease-related malnutrition with inflammation). On the other hand, predictive validity (eg, 30-day mortality in acute hospitalized patients) may be better determined by one of the phenotypic variables (eg, muscle parameters) in combination with inflammation.

Muscle mass is increasingly recognized as an important predictor of negative health outcomes and is now included in the GLIM diagnostic criteria. This is a great step forward as compared with previous operational definitions for malnutrition, which were mostly based on weight loss and BMI. The GLIM paper provides different options to identify a low muscle mass, whereby choices may depend on resources, time, and availability of specific cut points. For now, this seems like the best way to move forward and to start using GLIM in practice. However, it is acknowledged that different ways to measure muscle mass provide different results. To give an example: In a cohort of patients with cancer undergoing 3 different measurements of muscle mass (by bioelectrical impedance, mid-upper arm muscle

circumference, and computerized tomography scan analysis), the prevalence of low muscle mass varied from 13% (upper arm muscle circumference) to 93% (bioelectrical impedance) in the same cohort.[36] During the next few years, there will be opportunities to optimize the GLIM criteria by considering optimal cut points for the different muscle mass measurements according to different ethnicities, taking into account the suggestions for good-quality validation studies as described here. In addition, muscle quality (not yet incorporated in GLIM) could be studied, as there is increasing evidence that not only muscle quantity but also muscle quality[37] and strength[38] are associated with clinical outcomes. These are joint tasks with the sarcopenia community.

Conclusion and Recommendations

The GLIM criteria for malnutrition are unique in that they, for the first time, provide a globally accepted starting point to categorizing patients as malnourished. This is going to aid our understanding of the magnitude of malnutrition across different continents and across different healthcare settings. For now, having this worldwide accepted diagnostic framework is a step forward, and we encourage the use of this framework to show how GLIM can influence decision making and patient treatment and outcome. Reporting on this use will be important for understanding the utility of GLIM in practice. Meanwhile, validation studies are necessary to work toward refinement of the GLIM criteria. This paper provides guidance for performing good-quality validation studies for nutrition measures overall and GLIM in particular. We call for authors to publish these validation papers in leading clinical nutrition journals (eg, *JPEN*, *Clinical Nutrition*, *JAND*). Submission of the protocols for these studies to the GLIM Working Group would help to track ongoing prospective studies. Next to individual studies, we recommend the building of a large database to merge data and to allow for determining cut points of various measures upon which GLIM is based for different (sub)populations and healthcare settings to further support the refinement of the GLIM criteria.

Statement of Authorship

H. Keller and M. A. E. de van der Schueren equally contributed to the conception and design of the research; H. Keller and M. A. E. de van der Schueren contributed to the design of the research; H. Keller and M. A. E. de van der Schueren contributed to the acquisition and analysis of the data; H. Keller and M. A. E. de van der Schueren contributed to the interpretation of the data; and H. Keller and M. A. E. de van der Schueren drafted the manuscript. All authors critically revised the manuscript, agree to be fully accountable for ensuring the integrity and accuracy of the work, and read and approved the final manuscript.

References

1. Soeters P, Bozzetti F, Cynober L, Forbes A, Shenkin A, Sobotka L. Defining malnutrition: a plea to rethink. *Clin Nutr*. 2017;36(3):896-901.
2. Soeters PB, Reijven PL, van Bokhorst-de van der Schueren MA, et al. A rational approach to nutritional assessment. *Clin Nutr*. 2008;27(5):706-716.
3. Cederholm T, Barazzoni R, Austin P, et al. ESPEN guidelines on definitions and terminology of clinical nutrition. *Clin Nutr*. 2017;36(1):49-64.
4. White JV, Guenter P, Jensen G, Malone A, Schofield M. Consensus statement of the Academy of Nutrition and Dietetics/American Society for Parenteral and Enteral Nutrition: characteristics recommended for the identification and documentation of adult malnutrition (undernutrition). *J Acad Nutr Diet*. 2012;112(5):730-738.
5. Scheutz P, Fehr R, Baechli V, et al. Individualised nutritional support in medical inpatients at nutritional risk: a randomised clinical trial. *Lancet*. 2019;393(10188):2312-2321.
6. Weijs PJM, Looijaard WGPM, Beishuizen A, Birbes ARJ, Oudemans-van Straaten HM. Early high protein intake is associated with low morality and energy overfeeding with high mortality in non-septic mechanically ventilated critically ill patients. *Critical Care*. 2014;18(6):701:1-10.
7. Jensen GL, Mirtallo J, Compher C, et al. Adult starvation and disease-related malnutrition: a proposal for etiology-based diagnosis in the clinical practice setting from the International Consensus Guideline Committee. *JPEN J Parenter Enteral Nutr*. 2010;34(2):156-159.
8. Cederholm T, Jensen GL, Correia M, et al. GLIM criteria for the diagnosis of malnutrition—a consensus report from the global clinical nutrition community. *Clin Nutr*. 2019;38(1):1-9.
9. Jensen GL, Cederholm T, Correia M, et al. GLIM Criteria for the Diagnosis of Malnutrition: A Consensus Report From the Global Clinical Nutrition Community. *JPEN J Parenter Enteral Nutr*. 2019;43(1):32-40.
10. McKhann GM, Knopman DS, Chertkow H, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement*. 2011;7(3):263-269.
11. Lublin FD, Reingold SC, Cohen JA, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology*. 2014;83(3):278-286.
12. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. *Age Ageing*. 2019;48(1):16-31.
13. Gibson RS. *Principles of nutrition assessment*. New York: Oxford University Press; 2005.
14. Baker JP, Detsky AS, Wesson DE, et al. Nutritional assessment: a comparison of clinical judgement and objective measurements. *N Engl J Med*. 1982;306(16):969-972.
15. Guigoz Y, Vellas B, Garry PJ. Mini Nutritional Assessment: a practical assessment tool for grading the nutritional state of elderly patients. *Facts Res Gerontol*. 1994;Supplement: Nutrition:15-58.
16. Jager-Wittenaar H, Ottery FD. Assessing nutritional status in cancer: role of the Patient-Generated Subjective Global Assessment. *Curr Opin Clin Nutr Metab Care*. 2017;20(5):322-9.
17. Bauer J, Ash S, Davidson WL, et al. Evidence based practice guidelines for the nutritional management of cancer cachexia. *Nutr Diet*. 2006;63(suppl. 2):S3-S32.
18. Power L, de van der Schueren MAE, Leij-Halfwerk S, et al. Development and application of a scoring system to rate malnutrition screening tools used in older adults in community and healthcare settings—a MaNuEL study. *Clin Nutr*. 2019;38(4):1807-1819.

19. Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*. 2012;344:e686.
20. Streiner D, Norman G. *Health Measurement Scales: A Practical Guide to Their Development and Use*. 2nd ed. New York: Oxford University Press; 1996.
21. Bannigan K, Watson R. Reliability and validity in a nutshell. *J Clin Nurs*. 2009;18(23):3237-3243.
22. Jones JM. Validity of nutritional screening and assessment tools. *Nutrition*. 2004;20(3):312-317.
23. van Bokhorst-de van der Schueren MA, Guaitoli PR, Jansma EP, de Vet HC. Nutrition screening tools: does one size fit all? A systematic review of screening tools for the hospital setting. *Clinical Nutr*. 2014;33(1):39-58.
24. Keller HH, Goy R, Kane SL. Validity and reliability of SCREEN II (Seniors in the community: risk evaluation for eating and nutrition, Version II). *Eur J Clin Nutr.* 2005;59(10):1149-1157.
25. Wierdsma N, Kruizenga H, Stratton R. *Dietetic Pocket Guide, Adults*. Amsterdam: Press VU; 2017.
26. Walsh T. Fuzzy gold standards: Approaches to handling an imperfect reference standard. *J Dentistry.* 2018;74(suppl 1):S47-S49.
27. Jones JM. Reliability of nutritional screening and assessment tools. *Nutrition*. 2004;20(3):307-11.
28. Deutz NEP, Ashurst I, Ballesteros MD, et al. The underappreciated role of low muscle mass in the management of malnutrition. *J Am Med Dir Assoc*. 2019;20(1):22-27.
29. Power L, Mullally D, Gibney ER, et al. A review of the validity of malnutrition screening tools used in older adults in community and healthcare settings—a MaNuEL study. *Clin Nutr ESPEN*. 2018;24: 1-13.
30. Cobb AN, Daungjaiboon W, Brownlee SA, et al. Seeing the forest beyond the trees: Predicting survival in burn patients with machine learning. *Am J Surg*. 2018;215(3):411-416.
31. Shrout PE. Measurement reliability and agreement in psychiatry. *Stat Methods Med Res*. 1998;7(3):301-317.
32. Henrique JR, Rodrigues CN, Ferreira ÁRS, Correia MITD. GLIM in practice: sensibility and prognostic value for the diagnosis of malnutrition of gastrointestinal surgical patients. *Clin Nutr*. 2019;38(1): S239.
33. van Rijssen NM, Rojer AGM, Trappenburg MC, et al. Is being malnourished according to the ESPEN definition for malnutrition associated with clinically relevant outcome measures in geriatric outpatients? *Eur Geriatr Med*. 2018;9(3):389-394.
34. Charney P. Nutrition screening vs nutrition assessment: how do they differ? *Nutr Clin Pract.* 2008;23(4):366-372.
35. Sullivan GM, Feinn R. Using effect size-or why the P value is not enough. *J Grad Med Educ*. 2012;4(3):279-282.
36. Blauwhoff-Buskermolen S, Langius JAE, Becker A, Verheul HMW, de van der Schueren MAE. The influence of different muscle mass measurements on the diagnosis of cancer cachexia. *J Cachexia Sarcopenia Muscle*. 2017;8(4):615-622.
37. Martin L, Hopkins J, Malietzis G, et al. Assessment of computed tomography (CT)-defined muscle and adipose tissue features in relation to short-term outcomes after elective surgery for colorectal cancer: a multicenter approach. *Ann Surg Oncol*. 2018;25(9):2669-2680.
38. Van Ancum JM, Scheerman K, Jonkman NH, et al. Change in muscle strength and muscle mass in older hospitalized patients: a systematic review and meta-analysis. *Exp Gerontol*. 2017;92: 34-41.