

How I assess comorbidities before hematopoietic cell transplantation

Mohamed L. Sorrow

Clinical Research Division, Fred Hutchinson Cancer Research Center, and Division of Medical Oncology, Department of Medicine, University of Washington School of Medicine, Seattle, WA

The hematopoietic cell transplantation-comorbidity index (HCT-CI) is a comorbidity tool suited for recipients of HCT. The index has been shown to sensitively capture the prevalence and magnitude of severity of various organ impairments before HCT and to provide valuable prognostic information after HCT. Many investigators have validated the discriminative power of the HCT-CI, but others have not. One concern is the consistency in comorbidity coding across different

evaluators, particularly in view of the relatively recent addition of the HCT-CI to the transplant evaluation process. In this article, comorbidity scoring was tested across different evaluators, and only a fair interobserver agreement rate could be detected. To address these issues, a brief training program is proposed here, consisting of systematic methodology for data acquisition and consistent guidelines for comorbidity coding that were summarized in a Web-

based calculator. In a validation patient cohort, this training program was shown to improve the interevaluator agreement on HCT-CI scores to an excellent rate with weighted κ values in the range of 0.89 to 0.97. This proposed training program will facilitate reliable assessment of comorbidities in the clinic and for research studies leading to standardization of the use of comorbidities in prediction of HCT outcomes. (*Blood*. 2013;121(15):2854-2863)

Introduction

Organ dysfunctions (comorbidities) were found to be associated with the outcome of treatment of a given primary disease^{1,2} and, in particular, cancer.^{3,4} In 2005, a hematopoietic cell transplantation-comorbidity index (HCT-CI) was introduced as a measure of organ dysfunctions that was suited for recipients of HCT.⁵ The HCT-CI was developed from the historical Charlson comorbidity index⁶ after introducing 3 conceptual changes: the use of laboratory and organ function tests to redefine pulmonary, hepatic, cardiac, and renal comorbidities; the inclusion of all comorbidities encountered in a cohort of HCT recipients at a single institution; and the estimation of new adjusted hazard ratios for the associations between comorbidities and nonrelapse mortality after HCT. These adjusted hazard ratios were then converted into weights that could be summated into a total score.

In validation cohorts of recipients of allogeneic HCT from 2 different institutions, the HCT-CI was demonstrated to have higher discriminative power than the Charlson comorbidity index, both for non-relapse mortality and overall survival.^{5,7} Many investigators reported on the valid association between HCT-CI scores and mortality in their respective single-institutions,⁸⁻¹⁷ whereas a few others disagreed.¹⁸⁻²² A discussion of the possible reasons for the lack of complete agreement by investigators on the validity of the HCT-CI is outside the scope of this article. Instead, the focus of this article is a single concern that is related to the degree of consistency in assigning comorbidity scores among evaluators. For example, a recent study reported a noticeably higher prevalence of comorbidities compared with other reports.^{22,23} As investigators continue to explore the validity of the HCT-CI and to use it in decision-making and prognostication studies, an urgent need has emerged to standardize the methods and guidelines for comorbidity evaluation. A valid and reliable system for comorbidity evaluation would not only ensure the calculation of an accurate total comorbidity score

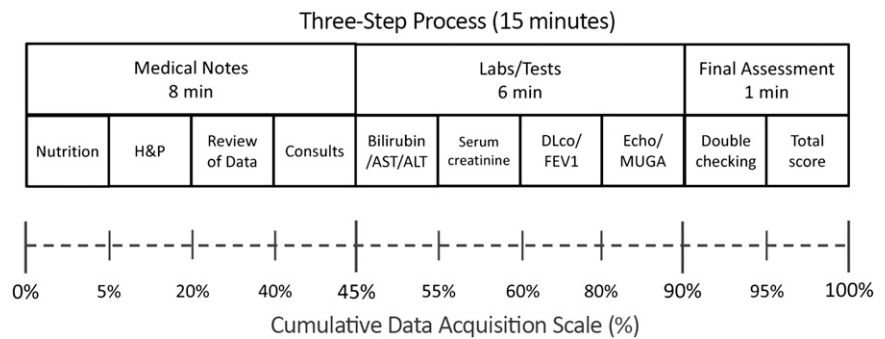
but also allow the accurate estimation of the prevalence of individual comorbidities, which would be of prime importance in future research addressing roles of comorbidities in post-HCT complications. Here, a brief training program is proposed comprising consistent methods for data acquisition from medical records and detailed guidelines for comorbidity assessment that were summarized in a Web-based application and a calculator. Validation of the ability of the proposed training program to improve the interrater reliability (IRR) of the HCT-CI is also described.

Training program

I. Methods of retrospective assessment of medical records for the acquisition of comorbidity data

Transplant physicians and physician assistants are conceivably familiarized with the use of well-established measures such as the Karnofsky scale for the assessment of performance status²⁴ and the systems used for grading acute GVHD.^{25,26} However, the introduction of new evaluation scales, even for a familiar clinical condition (eg, chronic GVHD), often requires the development of systematic methodology that could ensure a stepwise pattern of assessment.²⁷ Such methodology should not only save time by avoiding back-tracking and duplication of efforts but would also endorse the consistency in the acquisition of data and the evaluation of a given medical condition. The HCT-CI is a distinct assessment tool that is relatively new in the transplant field. Clinicians and study coordinators with varying degrees of experience could complete

Figure 1. Schema for the 3-step methodology for acquisition of comorbidity data from medical records.



a comorbidity evaluation in approximately 15 minutes by following the proposed 3-step evaluation process (Figure 1).

The landmark date is the date before HCT that will not be exceeded for any of the comorbidity-related evaluations. Day -10 was chosen to be the landmark date because all the conditioning regimens for patients who contributed to the development of the HCT-CI⁵ started after that day. As long as medical events and laboratory or organ function tests were done on or before the landmark date (day -10), they are suitable for the purpose of comorbidity coding. The use of a fixed time point (landmark date) for all evaluations would facilitate the retrieval of laboratory data from computer databases and the use of a Web-based calculator and would also standardize data collection across institutions. In the circumstance that a patient is given a conditioning regimen that starts before day -10, an evaluator could use the day before the start of conditioning regimen as the landmark date for assessment of comorbidities in that patient.

The following are the steps for the evaluation process:

1. Review of the important sections of the medical records:
 - a. Evaluate the nutrition notes to capture the measures of weight and height that were assessed at the closest time before the landmark date and then calculate the body mass index (BMI).
 - b. Assess the history and physical examination (H&P) note (the note dictated by the transplant physician or physician assistant within a few weeks before the landmark date). This note should include evaluations of the patient's present, past, social, and family history; a review of organ systems; and a physical examination of the patient. An evaluator should pay close attention to 3 major parts in the H&P note:
 - i. the past medical history for all details of recent and remote organ dysfunctions;
 - ii. the current medication list, to aid in the evaluation of some comorbidities and to detect others that might have been inadvertently dropped from the past medical history (eg, antidepressant medications for depression or oral hypoglycemic drugs for diabetes mellitus); and
 - iii. the final assessment summary for additional details on organ dysfunctions and for information on any planned consults/evaluations.

The H&P note could also be used as a good source for data on other important prognostic variables such as prior treatment and performance status scores. If the H&P note is exceptionally deficient in details on specific comorbidities, an evaluator should search for and review any prior notes on organ-specific problems (eg, gastroenterology consult notes on inflammatory bowel disease or a previous H&P summarized by the patient's primary oncologist or general medical practitioner) to confirm the diagnosis of

a given comorbidity and to determine whether a specific treatment was given or not.

- c. Examine the note summarized by the transplant physician that describes the findings from the pretransplant evaluation. In some institutions, this note would also include details on consenting for clinical trials. This note could be referred to as the "review of data" note. Other institutions might have this document in a different format. For example, findings of the pretransplant evaluations could be summarized in an updated H&P note before the start of conditioning regimen. It is anticipated that the pretransplant evaluation period generally spans 2 to 3 weeks before the start of the conditioning regimen. The "review of data" note, most frequently, is an abundant source for:
 - i. the most recent laboratory data,
 - ii. organ function tests,
 - iii. finalized recommendations from any requested consults, and
 - iv. current status and staging of the primary disease.
- d. Review notes on any requested consults (eg, a psychiatric consult for assessment of depression or anxiety) during the pretransplant evaluation period for:
 - i. assessment of severity of a given comorbidity and
 - ii. any recommended treatment specific for a given comorbidity.
2. Review of laboratory and organ function tests:
 - a. Assess the most updated report of pulmonary function test (PFT) before the landmark date. The PFT report contains details on:
 - i. the percentage of measured-to-predicted forced expiratory volume in 1 second (FEV1; Figure 2) and
 - ii. the percentage of measured-to-predicted diffusion capacity of carbon monoxide (DLco) after correction for hemoglobin (Figure 2).
 - b. Evaluate the echocardiogram or the multigated acquisition scan report for:
 - i. the percentage of ejection fraction (EF) for adults or shortening fraction (SF) for children,
 - ii. details on the presence and magnitude of severity of any valve abnormality, and
 - iii. details on other cardiac comorbidity (eg, dilated cardiomyopathy).
 - c. Assess liver function tests between days -24 and -10 (or between days -40 and -10 if only a single value is reported between days -24 and -10) before HCT for values (Figure 3) of:
 - i. alanine aminotransferase (ALT),
 - ii. aspartate aminotransferase (AST), and

		Ref	Pre Meas	Pre % Ref
A				
<u>Spirometry</u>				
FEV1	Liters	2.50	2.42	97
<u>Diffusion</u> Hb: 11.1 gm/dL				
DLCO	mL/mmHg/min	24.2	15.1	63
DLCocorr			19.6	81%
B				
<u>Spirometry</u>				
FEV1	Liters	4.11	3.17	77
<u>Diffusion</u> Hb: 11.0 gm/dL				
DLCO	mL/mmHg/min	38.1	22.7	60
DLCocorr			29.7	78%
C				
<u>Spirometry</u>				
FEV1	Liters	3.82	2.70	71
<u>Diffusion</u> Hb: 11.5 gm/dL				
DLCO	mL/mmHg/min	35.6	14.8	41
DLCocorr			18.4	52%
D				
<u>Spirometry</u>				
FEV1	Liters	3.04	1.44	47
<u>Diffusion</u> Hb: 10.4 gm/dL				
DLCO	mL/mmHg/min	28.0	11.3	40
DLCocorr			15.6	56%

Figure 2. PFTs for 4 different patients. There are 4 reports of PFTs for 4 different patients. Each report contains 5 columns of information in the following order: type/name of the test, unit of the test, reference "predicted" value for each test, prebronchodilator-measured value for each test, and prebronchodilator percentage (measured divided by predicted values) for each test. Postbronchodilator values and percentages were not shown here, as they do not contribute to the coding of pulmonary comorbidity per the HCT-CI. All uncorrected values or percentages of measured DLco were corrected for the concurrent hemoglobin values, using the Dinakara equation [corrected DLco = uncorrected DLco/(0.06965 × hemoglobin g/dL)]. (A) This is a patient who had an FEV1 percentage of 97%. The patient also had an uncorrected DLco percentage of 63%. The patient's concurrent hemoglobin value was 11.1 mg/dL; hence, he had a corrected DLco percentage of 81%. This patient was assigned a score of 0 for pulmonary comorbidity. (B) This patient had an FEV1 percentage of 77% and also experienced an uncorrected DLco percentage of 60%. Because the concurrent hemoglobin value was 11.0 mg/dL, the corrected percentage of DLco is 78%. This patient was assigned a score of 2 for pulmonary comorbidity because of both the FEV1 percentage of 77% and the corrected DLco percentage of 78%. (C) This is a patient with an FEV1 percentage of 71%. The uncorrected DLco percentage was 41%. The concurrent hemoglobin value was 11.5 mg/dL; therefore, the patient has a corrected DLco percentage of 52%. This patient was assigned a score of 3 for pulmonary comorbidity because of the corrected DLco percentage of 52%. (D) This patient experienced an FEV1 percentage of 47%. The uncorrected DLco percentage was 40%. The concurrent hemoglobin value was 10.4 mg/dL; therefore, the corrected DLco percentage was 56%. This patient was assigned a score of 3 for pulmonary comorbidity because of both the FEV1 percentage of 47% and the corrected DLco percentage of 56%.

- iii. total bilirubin.
- d. Assess serum creatinine values between days -24 and -10 (or between days -40 and -10 if only a single value is reported between days -24 and -10) before HCT.
3. Summary and final assessment:
 - a. Double checking:
 - i. As an evaluator goes through the different sections of the medical record, it would be helpful to enter each

positive finding momentarily in a software spreadsheet tool such as MS Excel, a Web-based calculator, or simply by pen and paper.

- ii. Once all 8 subsections of the 2 major components of data acquisition (Figure 1) are completed,
 1. double-check all the positive findings listed in the calculator or the sheet and
 2. fix any incorrect data entry that does not fit the patient's overall presentation, as at that time an evaluator would have recent recollection of details of the medical record and quick access to the chart to verify any information.
- b. Calculate a total score and assign it to the patient chart.

II. Guidelines for assessment of comorbidities in the HCT-CI (the Comorbidity Coding Tool)

All clinical and laboratory criteria described in this coding tool are meant for the evaluation of comorbidities specifically per the HCT-CI (Table 1).

1. Arrhythmia (score 1). A score of 1 is assigned for any type of arrhythmia that has necessitated the delivery of a specific antiarrhythmia treatment at any time in the patient's past medical history. Examples include atrial fibrillation or flutter, sick sinus syndrome, or ventricular arrhythmias. A score is assigned even if the patient was in normal sinus rhythm at the time of data acquisition or at the landmark date. No score is assigned to transient arrhythmias that never required treatment.

Sometimes, the medical record does not include enough details on the treatment of a prior arrhythmia. In this case, judgment should be based on the clinical significance of the described arrhythmia and the clinical situation accompanying the development and resolution of such an arrhythmia. For example, a patient who developed a rapid atrial fibrillation requiring admission and management in the intensive care unit is assigned a score for arrhythmia even if the medical record does not state the type, dose, and duration of treatment given. However, if the clinical situation after careful review of the medical record raises doubt as to whether or not a treatment was given for an arrhythmia, no score is assigned. An example could be a patient who once developed a paroxysmal atrial fibrillation with no indication of a rapid ventricular response and that arrhythmia resolved spontaneously with no mention of the use of a specific antiarrhythmia treatment.

2. Cardiovascular Comorbidity (score 1). A maximal score of 1 is assigned for cardiovascular comorbidity in the presence of 1 or more of the following 3 clinical presentations.

Coronary artery disease. This is based on the presence of a documented diagnosis of chronic exertional angina, unstable angina, or myocardial infarction at any point in the patient's past medical history, as stated in the H&P section of the medical record. Information on prior placement of a coronary stent or undergoing a coronary artery bypass graft surgery should support coding this comorbidity.

Congestive heart failure. To score this clinical presentation, the medical record should have a statement about the development of symptoms/signs of congestive heart failure (eg, an exertional or paroxysmal nocturnal dyspnea) that later responded to diuretics, afterload-reducing agents, β blocker, and/or digitalis at any time in the patient's past medical history.

Low EF. Patients with an EF of 50% or lower or a SF (for pediatric patients) of 26% or lower as determined by an echocardiogram or a multigated acquisition scan are assigned a score of 1. As noted before, evaluation of this comorbidity item should be restricted

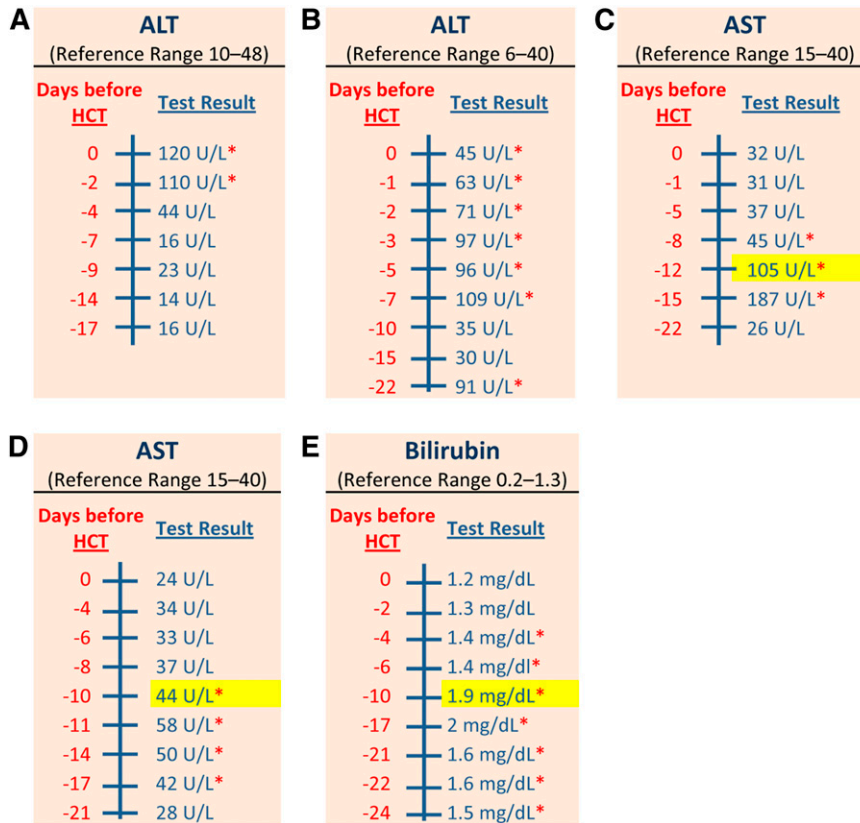


Figure 3. Hepatic function tests for 5 patients. There are 5 reports of hepatic function tests for 5 different patients. Each report contains 2 columns of information in the following order: a scale for days in relation to the HCT and the measured values for each test. Name and reference range for each test is provided in the top row of each report. Laboratory values between days –24 and –10 before HCT were the only values considered for evaluation of hepatic comorbidity. The elevated laboratory value that was closest to day –10 was used for assigning the severity score; it was highlighted in yellow. The red asterisk (*) means values were outside the reference range. (A) This patient had 2 normal values of ALT at days –17 and –14 before HCT. No score was assigned to the ALT portion of hepatic comorbidity for this patient. (B) This patient had only 3 values of ALT between days –24 and –10 before HCT; 2 values were normal and the other value was elevated. Even though this patient had multiple elevated values of ALT after day –10, no score was assigned for the ALT portion of hepatic comorbidity because the patient did not have 2 elevated values of ALT between days –24 and –10. (C) This patient had 1 normal and 2 elevated values of AST between days –22 and –12 before HCT. The maximal score, a score of 3, was assigned for hepatic comorbidity in this patient because the closest elevated value of AST to day –10 was 105 U/L, which was more than 2.5 times the ULN for AST based on the reference range. Note that the values after day –12 did not contribute to the evaluation of this comorbidity. (D) This patient had 1 normal and 4 elevated values of AST between days –21 and –10 before HCT. A score of 1 was assigned to the AST portion for hepatic comorbidity in this patient, given that the closest elevated value of AST to day –10 was 44 U/L, which was less than 2.5 the ULN for AST. A score of 1 was assigned despite normalization of AST values after day –10. A complete evaluation of other hepatic laboratory tests (ALT and bilirubin) was still required before a decision is made on the maximal score to be assigned for hepatic comorbidity in this patient. (E) This patient had multiple elevated values of total bilirubin between days –24 and –10 before HCT. A score of 1 was assigned for the bilirubin portion of hepatic comorbidity in this patient because the value of total bilirubin at day –10 was 1.9 mg/dL (<2.5 the ULN for total bilirubin). A complete evaluation of other hepatic laboratory tests (ALT and AST) was required before a decision is made on the maximal score to be assigned for hepatic comorbidity in this patient.

to the most recent measurements of EF or SF before the landmark date. Lack of evaluation of EF or SF for an individual patient before transplant does not preclude the calculation of a total HCT-CI score for that patient.

3. Inflammatory bowel disease (score 1). A score of 1 is assigned for this comorbidity on the basis of the presence of a documented prior diagnosis (history of an endoscopic examination of the mucosa with or without confirmatory histology and radiologic findings) of Crohn’s disease or ulcerative colitis requiring treatment at any time in the patient’s past medical history. If the patient has never received a treatment of this comorbidity, no score is assigned.

4. Diabetes (score 1). A score of 1 is assigned for this comorbidity on the basis of the diagnosis of diabetes or steroid-induced hyperglycemia requiring continuous treatment with insulin or oral hypoglycemic drugs during the instantaneous period of 4 weeks before the landmark date. No score is assigned for this comorbidity if diabetes could be controlled with diet alone or if a previous treatment of diabetes or steroid-induced hyperglycemia was stopped 4 weeks before the landmark date.

5. Cerebrovascular disease (score 1). A score of 1 is assigned for cerebrovascular disease on the basis of a prior diagnosis of transient ischemic attack, subarachnoid hemorrhage, or cerebral thrombosis, embolism, or hemorrhage at any time in the past medical history. No details on treatment are required for assigning a score for this comorbidity.

6. Psychiatric disorder (score 1). A score of 1 is assigned for this comorbidity on the basis of the presence of any mood, anxiety, or other psychiatric disorder requiring continuous treatment during the instantaneous period of 4 weeks before the landmark date. Depression and anxiety are the most common psychiatric disorders encountered in transplant populations, yet other disorders such as schizophrenia or bipolar disorder should also be coded for this comorbidity. Patients who are receiving only “as-needed” medications for any of the above disorders are not assigned a score for this comorbidity.

7. Hepatic comorbidity (2 levels of severity). As a general rule, assessment of the laboratory tests (a and/or b) has to include at least 2 values per test on 2 different days within a period extending between

Table 1. HCT-CI

Comorbidities	HCT-CI scores
Arrhythmia	1
Cardiovascular comorbidity	1
Inflammatory bowel disease	1
Diabetes or steroid-induced hyperglycemia	1
Cerebrovascular disease	1
Psychiatric disorder	1
Mild hepatic comorbidity	1
Obesity	1
Infection	1
Rheumatologic comorbidity	2
Peptic ulcer	2
Renal comorbidity	2
Moderate pulmonary comorbidity	2
Prior malignancy	3
Heart valve disease	3
Moderate/severe hepatic comorbidity	3
Severe pulmonary comorbidity	3
Total score = _____	

days -24 and -10 before HCT (Figure 3). That period could be extended to be between days -40 and -10 only in the case that liver function tests were done only once between days -24 and -10 before HCT. The laboratory value closest to the landmark date should be the value used in defining the severity of hepatic comorbidity (Figure 3). The upper limit of normal (ULN) for any of the 3 tests is determined on the basis of the reference range per the institution laboratory.

A) Mild hepatic comorbidity (score 1): A maximal score of 1 is assigned for mild hepatic comorbidity in the presence of 1 or more of the following 3 clinical presentations: (1) elevated total bilirubin to a value higher than the ULN and up to 1.5 times the ULN; (2) elevated values of any of the 2 hepatic transaminase enzymes, ALT or AST, to values higher than the ULN and up to 2.5 times the ULN; or (3) a prior diagnosis of an infection with hepatitis B or C at any time in the patient's past medical history before the landmark date.

B) Moderate to severe hepatic comorbidity (score 3): A maximal score of 3 is assigned for moderate to severe hepatic comorbidity in the presence of 1 or more of the following 3 clinical presentations: (1) elevated values of total bilirubin to a level higher than 1.5 times the ULN; (2) elevated values of any or both of the 2 hepatic transaminase enzymes to levels higher than 2.5 times the ULN; or (3) a documented diagnosis of liver cirrhosis at any time in the patient's past medical history before the landmark date.

8. Obesity (score 1). A score of 1 is assigned for obesity based on a BMI higher than 35.00 kg/m^2 for patients older than 18 years or a BMI for age of the 95th percentile or higher for patients aged 18 years or younger. Evaluation of this comorbidity is based on the most recent measurement of the BMI (or weight and height needed for the calculation of the BMI) before the landmark date.

9. Infection (score 1). A maximal score of 1 is assigned for infection comorbidity in the presence of 1 or more of the following 4 clinical presentations: (1) a documented infection (eg, by culture or biopsy), (2) fever of unknown origin, (3) pulmonary nodules suspicious for fungal pneumonia, or (4) a positive purified protein derivative test requiring prophylaxis against tuberculosis. Patients must have started a specific antimicrobial treatment before the landmark date with a recommendation, documented in the chart either by the primary team or the infection consult team, to continue the same antimicrobial therapy (or a similar agent) during the days of administration of a conditioning regimen and beyond day 0 of HCT.

10. Rheumatologic comorbidity (score 2). A score of 2 is assigned for a rheumatic comorbidity on the basis of the presence of a documented prior diagnosis of a rheumatologic disease that has required administration of a specific treatment at any time in the patient's past medical history. Diagnoses include systemic rheumatologic and connective tissue disorders such as systemic lupus erythematosus, rheumatoid arthritis, Sjögren's syndrome, scleroderma, polymyositis, dermatomyositis, mixed connective tissues disease, polymyalgia rheumatica, polychondritis, sarcoidosis, and vasculitis syndromes. Patients with undiagnosed polyarthritis, degenerative joint disease, or osteoarthritis are not scored for this comorbidity. Occasionally, a patient might have a clinical pattern of a systemic rheumatologic disease responding to a specific treatment but without a definitive diagnosis. For example, I was consulted once on a patient with an unspecified collagen vascular disease that had presented 4 years earlier and had manifested by iritis, uveitis, bowel disturbances, and muscle aches. This unspecified collagen vascular disease responded to low-dose systemic steroids. Even though there was no definitive rheumatologic diagnosis in this case, I erred on the side of caution and assigned a score for this presentation as a rheumatologic comorbidity.

Patients with quiescent rheumatologic diseases who are receiving no treatment in the immediate period before the landmark date are assigned a score for this comorbidity if they have fulfilled the prior criteria.

11. Peptic ulcer (score 2). A score of 2 is assigned for peptic ulcer on the basis of the presence of a prior endoscopic or radiologic diagnosis of gastric or duodenal ulcer, noted in the medical record, at any point in the patient's past medical history. Patients with quiescent peptic ulcer who are receiving no treatment in the immediate period before the landmark date are assigned a score for this comorbidity if they have met the prior criteria.

12. Renal comorbidity (score 2). A maximal score of 2 is assigned for renal comorbidity in the presence of 1 or more of the following 3 clinical presentations: (1) elevated values of serum creatinine to more than 2 mg/dL or more than $176.8 \text{ } \mu\text{mol/L}$ (Figure 4), as detected in at least 2 laboratory tests on 2 different days within a period extending between days -24 and -10 before HCT (this evaluation period could be extended to span between days -40 and -10 if serum creatinine was evaluated only once between days -24 and -10 before HCT); (2) chronic renal disease requiring weekly dialysis within the instantaneous period of 4 weeks before the landmark date; or (3) a documented prior history of renal transplantation at any point in the patient's past medical history.

13. Pulmonary comorbidity (2 levels of severity). As a general rule, assessment of pulmonary comorbidity for the purpose of assigning HCT-CI scores should exclusively rely on PFT results, and in particular corrected DLco and FEV1 percentages (Figure 2). A total HCT-CI score should not be calculated in the absence of data on PFT except in the case that PFT could not be done because of technical difficulties (eg, in pediatric patients). Occasionally, patients are assessed by a postbronchodilator (reversibility) test. In this case, only the prebronchodilator values of FEV1 are considered for evaluation of pulmonary comorbidity.

Measured DLco values should first be corrected for the concurrent hemoglobin value using the Dinakara equation (Corrected DLco = uncorrected DLco/($0.06965 \times \text{hemoglobin g/dL}$)).²⁸ Then, the corrected value of measured DLco is divided by the predicted value to compute the percentage of DLco. Alternatively, the uncorrected DLco percentage, which is reported in all PFT reports, could be directly corrected for the concurrent hemoglobin value, using the Dinakara equation to compute the corrected DLco percentage (Figure 2). Either way will lead to the same final

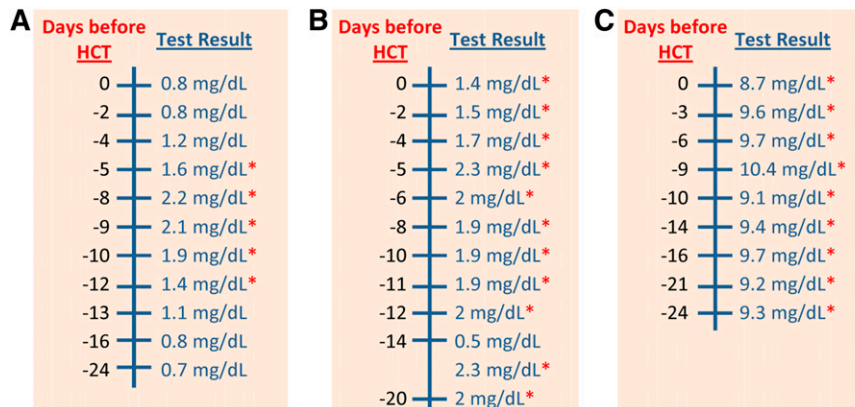


Figure 4. Serum creatinine values for 3 patients. There are 3 reports of serum creatinine for 3 different patients. Each report contains 2 columns of information in the following order: a scale for days in relation to the HCT and the measured values for each test. Reference range for each test is provided in the top row of each report. The red asterisk (*) means values were outside the reference range. The reference range for serum creatinine was 0.3 to 1.2 mg/dL. Values of serum creatinine that were tested between days -24 and -10 before HCT were the only values considered for evaluation of renal comorbidity. (A) This patient had 3 normal and 2 elevated values of serum creatinine between days -24 and -10 before HCT. No score was assigned for renal comorbidity in this patient because both of the elevated values were less than 2 mg/dL. The 2 values of serum creatinine at days -9 and -8, which were higher than 2 mg/dL, did not contribute to the evaluation of this comorbidity because they were assessed after day -10. (B) This patient had consistently elevated values of serum creatinine throughout the whole laboratory report. However, no score was assigned for renal comorbidity in this patient because there was only a single value of more than 2 mg/dL at day -14, which was not sufficient for coding this comorbidity. (C) This patient had consistently elevated values of serum creatinine throughout the whole laboratory report. A score of 2 was assigned for renal comorbidity in this patient because there were 2 values of serum creatinine higher than 2 mg/dL between days -24 and -10 before HCT.

percentage of corrected DLco. The Dinakara equation is favored over other equations such as the one by Cotes et al²⁹ because of its more robust ability to account for the effects of anemia, a common sign of the primary hematologic disease, and because it is the equation used by the PFT laboratory at Fred Hutchinson Cancer Research Center (FHCRC), where the HCT-CI was originally developed.

Moderate pulmonary comorbidity (score 2). A maximal score of 2 is assigned for moderate pulmonary comorbidity in the presence of 1 or more of the following 3 clinical presentations: (1) a percentage of DLco in the range of 66% to 80% or (2) a percentage of FEV1 in the range of 66% to 80% (both should be the most recent measurements before the landmark date), or (3) shortness of breath on slight activity that is attributed to a pulmonary disease and cannot be corrected by blood transfusion for a noticeable anemia, as assessed during a clinic visit within the immediate period of 2 weeks before the landmark date.

Severe pulmonary comorbidity (score 3). A maximal score of 3 is assigned for severe pulmonary comorbidity in the presence of 1 or more of the following 4 clinical presentations: (1) a percentage of DLco of 65% or less or (2) a percentage of FEV1 of 65% or less (both should be the most recent measurements before the landmark date); (3) shortness of breath at rest that is attributed to a pulmonary disease and cannot be corrected by blood transfusion for a noticeable anemia, as assessed during a clinic visit within the immediate period of 2 weeks before the landmark date; or (4) the need for intermittent or continuous oxygen supplementation during the immediate period of 4 weeks before the landmark date.

14. Prior malignancy (score 3). A score of 3 is assigned for this comorbidity on the basis of the presence of a prior diagnosis of any malignancy that required receiving a specific treatment at any point in the patient's past medical history, regardless of the type of treatment (surgery, radiotherapy, and/or drug therapy). Lymphomas or myelomas that preceded the diagnosis of a myeloid malignancy (eg, acute myeloid leukemia [AML], myelodysplastic syndromes, or chronic myeloid leukemia) are assigned a score for this comorbidity. Similarly, myeloid malignancies that preceded the diagnosis of lymphomas or myelomas are assigned a score for this comorbidity.

Patients with a prior malignancy from the same lineage of cells of the current malignancy should not be assigned a score for this comorbidity; for example, if a patient had a diagnosis of non-Hodgkin lymphoma that was preceded by Hodgkin lymphoma or if a patient had a diagnosis of AML that was preceded by myelodysplastic syndromes.

Melanoma, but not basal or squamous cell carcinoma of the skin, should be assigned a score for this comorbidity. Patients with a prior malignancy that never required a specific treatment are not scored for this comorbidity. Tumors of a benign nature are not scored for this comorbidity.

15. Heart valve disease (score 3). A maximal score of 3 is assigned for heart valve comorbidity in the presence of 1 or more of the following 3 clinical presentations: (1) at least a moderate or severe degree of valve stenosis or insufficiency, as determined by echocardiogram, whether that valve was mitral, aortic, tricuspid, or pulmonary; (2) prosthetic mitral or aortic valve; or (3) symptomatic mitral valve prolapse. Assessment of this comorbidity is limited to the most recent heart evaluation by echocardiogram before the landmark date.

Use of the guidelines for prospective assessment of comorbidities. The guidelines described in the previous section were meant for retrospective evaluation of comorbidities when the patient has already passed the landmark date of day -10. Retrospective evaluation of comorbidities could be used for prognostic studies or in comparative effectiveness research about the HCT. The HCT-CI could also be evaluated prospectively by clinicians or study coordinators before day -10 for the purpose of risk-benefit assessment; for example, at the time of a transplantation consult to aid in the decision-making about the intensity of conditioning regimen. In that situation, the previous guidelines would apply with a change in the landmark date to be the date of the consult. Similarly, investigators assessing comorbidities using the HCT-CI before conventional therapeutic interventions, such as induction chemotherapy for AML,^{20,21} could use the date of comorbidity assessment as the landmark date.

For example, a transplant physician is seeing a patient for a consult on December 30, 2012. In this case, the landmark date will be December 30, 2012. The physician should use the values of

Table 2. Evaluation of the IRR rates for assignment of the HCT-CI scores

Study phases	Evaluators	Patients, n	Agreement on individual scores, %	Kw* (standard error)
Initial assessment phase	Principal investigator vs evaluator 1	80	38	0.585 (0.114)
	Principal investigator vs multiple evaluators	80	37	0.552 (0.065)
	Evaluator 1 vs multiple evaluators	80	38	0.433 (0.076)
Initial validation phase	Principal investigator vs evaluator 1	90	78	0.910 (0.03)
	Principal investigator vs evaluator 2	90	78	0.890 (0.03)
	Principal investigator vs evaluator 3	90	70	0.910 (0.03)
Final validation phase	Principal investigator vs evaluator 4	30	83	0.970 (0.03)
	Evaluator 1 vs evaluator 4	30	77	0.920 (0.05)
	Evaluator 2 vs evaluator 4	30	77	0.910 (0.05)
	Evaluator 3 vs evaluator 4	30	79	0.930 (0.05)

*Weighted κ statistic (Kw) was calculated for agreement on HCT-CI risk groups of 0 to 1, 2, 3, and 4 or more.

hepatic function tests that are done between December 16, 2012, and December 30, 2012, to assess hepatic comorbidity. If only a single value of bilirubin is available during this interval, then the assessment period can be extended to be between December 1, 2012, and December 30, 2012. In addition, this patient will have to be continuously treated with an antidiabetic or a psychiatric treatment for 4 weeks between December 3, 2012, and December 30, 2012, for a score to be assigned for diabetes or psychiatric comorbidity, respectively. If the patient has a diagnosis of an infection before the consult date, it can only be scored as a comorbidity if the prescribed antimicrobial medication is required to be continued for more than 10 days (until January 10, 2013) after the date of the consult.

III. The Web-based HCT-CI score calculator

The explanatory guidelines provided here were adapted into a Web-based application and calculator, available at <http://www.hctci.org>. Evaluators are assigned a registered, password-protected access to the Web site, where they can save portions of a patient's comorbidity data as they become available in the clinic or during chart review, under a de-identified patient-specific code. Evaluators can access stored data until all comorbidity data are collected and a total score is requested. The calculator contains 15 categories of comorbidities per the HCT-CI (pulmonary and hepatic comorbidities entail 2 grades of severity), and under each category there are several choices for different clinical presentations. The evaluator is requested to enter the following information: date of the transplant; measures of weight and height to calculate the BMI; 2 values for each of the laboratory tests for AST, ALT, bilirubin, and creatinine with the corresponding date for each value; and percentages for EF, SF, FEV1, and uncorrected DLco with the concurrent hemoglobin value.

Options are available if the PFT was not done because of younger age, if EF or SF were not done, or if an extended evaluation period (between days -40 and -10) is needed for some laboratory tests. In this situation, the Web-based application will perform the following actions:

- it will calculate the BMI and determine whether or not a score should be assigned for obesity;
- it will provide the corrected DLco percentage and will determine, based on percentages of both DLco and FEV1, whether or not and which score (1 or 3) should be assigned for pulmonary comorbidity;
- it will determine the score to be assigned for hepatic and renal comorbidities based on the laboratory values and their dates; and
- it will assign scores for other comorbidities based on the selected clinical situations. Finally, a total score could be generated. The

Web-based application will also provide a summary of all positive comorbidities for a given patient.

If a patient is being evaluated prospectively for comorbidities at a stage preceding the pretransplant evaluation period (eg, during an early consult for HCT), then the evaluator should substitute the date of transplant in the Web-based application with a hypothetical date that is 10 days after the date of the consult. Finally, the Web-based calculator was tested and validated several times by the principal investigator (PI) and the comorbidity evaluation team.

Assessment of the IRR rates for validation of the proposed training program

Patients and methods

Assessment of the IRR was done using data from randomly selected samples of patients who received allogeneic or autologous HCT at FHCRC. This retrospective study was approved by the Internal Review Board of FHCRC and conducted in accordance with the Declaration of Helsinki. Other than the PI (M.L.S.), none of the evaluators had any prior experience in evaluating comorbidities, and they had either limited or no experience in HCT.

The assessment of the IRR was done over the course of 3 phases.

Initial assessment phase. A sample of 88 patients was randomly selected for comorbidity evaluation during this phase. The PI and another evaluator (evaluator 1) independently collected comorbidity data from medical records of the 88 patients and then assigned the HCT-CI scores. Evaluator 1 was a first-year fellow in the Hematology-Oncology Program and used the HCT-CI as it was previously published,⁵ with no further assistance in comorbidity coding from the PI. Then, the HCT-CI scores from the PI and evaluator 1 were forwarded to the biostatistician for comparison. In addition, scores assigned by both single evaluators (PI and evaluator 1) were compared with those previously determined by multiple evaluators in the clinic (Table 2). The "multiple evaluators" were the medical providers who took care of the 88 patients while receiving their transplant and evaluated their comorbidities prospectively.

Initial validation phase. As described earlier, the brief training program was developed to achieve substantial agreement on comorbidity coding by different evaluators. Three evaluators contributed to this phase: evaluator 1, who contributed to the previous phase, and 2 other novice evaluators, evaluator 2 and evaluator 3.

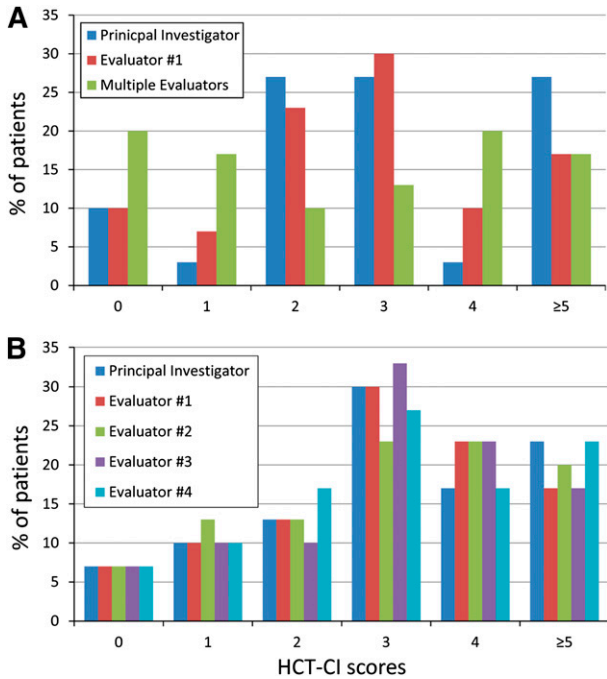


Figure 5. The frequency of distribution of the raw scores for the HCT-CI by different evaluators. (A) There was a great variation in the frequency of assigning the raw scores of 0, 1, 2, 3, 4, and 5 or more by any of the 3 groups of evaluators during the initial assessment phase. (B) A limited variation could be seen in the frequency of assigning the raw scores of 0, 1, 2, 3, 4, and 5 or more by the 5 evaluators during the final validation phase.

Evaluators 2 and 3 were graduates of foreign medical schools with no prior clinical or research experience in the United States. An additional sample of 98 patient charts was randomly selected for this phase. The PI printed out the HCT-CI and the documents for the new training program and handed them to the 3 evaluators. The PI held 60-minute-long sessions with each evaluator to review the steps for accessing the medical records for data acquisition and to answer any questions about the comorbidity coding tool. Then the PI and the 3 evaluators independently collected the comorbidity data from the medical records of the 98 patients and assigned the HCT-CI scores. The scores were then forwarded to the biostatistician. This phase had 2 aims: to demonstrate an improvement in the IRR of evaluator 1 compared with that in the initial phase, and to show that novice evaluators, evaluators 2 and 3, could demonstrate excellent IRR rates when provided firsthand with the proposed training program.

Final validation phase. The Web-based application and calculator were established, including the guidelines for scoring each of the 17 comorbidities. A fourth evaluator, evaluator 4, was recruited to validate the training program and the Web-based application. Evaluator 4 was a first-year medical student at the University of Washington. Among the 98 patients' charts included in the initial validation phase, a sample of 30 patient charts was randomly selected for the final validation phase. The PI handled the documents for the training program, including the Web site for the Web-based calculator to evaluator 4. Evaluator 4 independently assigned scores to the 30 charts, and the scores were compared with those previously determined by the PI and each of the other 3 evaluators during the initial validation phase.

Statistical methods

The κ statistic is a measure used to analyze interrater agreement.^{30,31} It adjusts for the degree of agreement that would be expected to occur

by chance and is therefore more appropriate than Pearson's product moment, Spearman's correlation, or percentage agreement.³² It is reported from 0.0 to 1.0. Weighted κ statistic (Kw),³³ which assigns less weight to agreement as risk categories are further apart, was computed with Fleiss-Cohen weights³⁴ to analyze the magnitude of interrater agreement between 2 raters on assignment of patients to the HCT-CI risk categories of 0 to 1, 2, 3, and 4 or more. Standard errors (SEs) for κ and Kw statistics were calculated as previously described.³⁵ The κ statistic could be used to assess the reliability of agreement between either 2 raters (Cohen's κ)³⁰ or multiple raters (Fleiss' κ),³¹ whereas weighted Kw is reserved for comparisons between 2 raters.³³ Although we report here on results using both methods of assessment, we were more interested in comparing the individual results between 2 raters, using the Kw, than the average scores among multiple raters, using Fleiss' κ statistic. The Landis scale was used for interpretation of the magnitude of κ and Kw statistics where values of 0 indicate no agreement, 0.01 to 0.20 indicate slight agreement, 0.21 to 0.40 indicate fair agreement, 0.40 to 0.60 indicate moderate agreement, 0.61 to 0.80 indicate substantial agreement, and 0.81 to 1.00 indicate almost perfect agreement.³⁶ During the initial phase, we considered that a Kw value below 0.60, although acceptable in some settings, would indicate the need to improve the IRR by new methods and guidelines for comorbidity evaluation. By developing the training program, our goal was to achieve a value of Kw greater than 0.80, indicating excellent agreement, among any 2 evaluators during the validation phase to validate the ability of the proposed training program to improve the interobserver congruence when assigning the HCT-CI scores.

Results

Initial assessment phase. Among the sample of 88 patients, evaluator 1 could assign scores for 80 patients, which was the final sample used for this phase. Figure 5A shows the magnitude of variations in the frequency of assigning raw scores when comparing the PI, evaluator 1, and the multiple evaluators. Variations existed among the 3 sets of evaluators and across all of the raw scores of 0, 1, 2, 3, 4, and 5 or more, but they were more pronounced when comparing the scores assigned by multiple evaluators versus each of the 2 single evaluators. The Fleiss' κ statistic (SE) for agreement on the average scores among the 3 groups of raters was 0.38 (0.06), indicating fair agreement. Likewise, among each 2 raters, the use of the HCT-CI to score comorbidities without instructive guidelines resulted in only fair interrater agreement, with Kw values ranging between 0.433 and 0.585 (Table 2). The Kw statistic was slightly better between the 2 single evaluators (0.585) compared with that between each of the 2 single evaluators and other multiple evaluators in the clinic (0.552 and 0.433, respectively). Overall, results indicated a modest rate of IRR without comorbidity coding methodology and guidelines.

Initial validation phase. Among the sample of 98 patient charts, evaluators 2, 3, and 4 could assign scores to 90 charts, which was the final sample for this phase. Evaluator 1 showed a substantial improvement in agreement on assigning the HCT-CI scores versus the PI, with a Kw (SE) of 0.91 (0.03). Similarly, the other 2 novice evaluators demonstrated excellent IRR rates, with a Kw of 0.89 to 0.91, with SEs of 0.03 for both.

Final validation phase. Evaluator 4 assigned scores to the total sample of 30 charts. Figure 5B showed a small magnitude of variation in the frequency of assigning the raw scores when comparing the 5 evaluators, and most of those variations were limited to the highest scores (4 and 5 or more). The Fleiss' κ statistic for agreement on the

average scores among the 5 evaluators was substantially improved to 0.80 (0.05) compared with the initial assessment phase. The Kw statistics among each group of 2 evaluators were all higher than 0.900, indicating almost perfect agreement (Table 2).

Summary

The HCT-CI can be used to capture the magnitude of organ damage before HCT for a given primary hematological disease. It is also an important tool for decision-making in the clinic, for comparative effectiveness studies of conditioning regimens and graft sources, and for adjustment of statistical analyses for prognostic studies. So far, the index has been evaluated in more than 25 publications from transplant centers worldwide. The index is expected to be continuously used in the HCT field. The Center of International Blood and Marrow Transplantation Research has incorporated the HCT-CI in routine data collection from transplant centers and will use the index and other variables in the Center Outcome Analyses designed to compare outcomes across transplant centers and to provide this information to patients, insurance companies, and academic investigators. In addition, studies on further refinements of the HCT-CI that improve prognostication but retain simplicity of the index would require large data from multiple institutions with consistently evaluated comorbidities. To achieve these goals, a consensus on comorbidity evaluation across centers is mandatory. Here, new methods and guidelines were proposed to facilitate consistent comorbidity coding. We have seen a fair degree of interobserver agreement when novices assessed comorbidities without standardized guidelines. Similar IRR rates are expected among evaluators from different institutions, given that comorbidity assessment is a newly introduced subspecialty to the transplant field.

It is important to report the IRR rates for comorbidity indices to ensure accurate comparison of results from clinical trials across institutions.³⁷ Multiple studies have previously reported variable IRR rates for several indices,³⁸⁻⁴⁰ yet little systematic effort has been made to improve the IRR for any comorbidity index. Here, efforts were made to enhance the agreement on the HCT-CI scoring by both improving the instrument (the comorbidity coding tool) and training the evaluators (the methodology and the Web-based application).^{32,41} The IRR was lowest (Kw, 0.433) when scores were compared between a single evaluator and multiple untrained evaluators in the clinic, suggesting the need for a training program to prepare experienced comorbidity evaluators at different institutions.

Participants in the current training program had no prior experience in comorbidity coding and limited or no prior experience in allogeneic HCT. Therefore, we expect that the proposed methods and guidelines could function appropriately in training a wide variety of individuals with different qualifications, ranging from study coordinators to experienced transplant physicians. Success of the training program was shown by improvements of both Fleiss' κ statistics, among multiple evaluators, from 0.380 to 0.800 and Kw values, among 2 evaluators, from between 0.433 and 0.585 to between 0.890 and 0.970 when comparing the initial versus the validation phases, respectively. We would expect these values to be maintained or improved upon when applied by center-specific transplant-oriented individuals. The Web-based HCT-CI includes a summary of all the explanatory guidelines for comorbidities and provides a user-friendly calculator of the scores. The methods, guidelines, and Web-based application together constitute a brief training program that could be used worldwide by evaluators at single institutions to standardize comorbidity coding.

Acknowledgments

The author thanks Dr Barry Storer for his help with the statistical part of IRR and Dr Fabiana Ostronoff, Saima Ijaz, Aisha Al-Khinji, and Jennifer McClure, for their participation in the assessment of interobserver agreement on comorbidity coding. The author also thanks Helen Crawford, Bonnie Larson, Sue Carbonneau, Joan Vermeulen, and Karen Carbonneau for their administrative assistance with study implementation and manuscript preparation.

This work was supported by a grant from the National Institutes of Health (HL088021).

Authorship

Contribution: M.L.S. designed, tested, and validated the training program and the Web-based calculator and wrote the paper.

Conflict-of-interest disclosure: The author declares no competing financial interests

Correspondence: Mohamed L. Sorror, Clinical Research Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, WA 98109-1024; e-mail: msorror@fhcrc.org.

References

- Feinstein AR. The pre-therapeutic classification of co-morbidity in chronic disease. *J Chronic Dis*. 1970;23(7):455-468.
- Fried LP, Bandeen-Roche K, Kasper JD, et al. Association of comorbidity with disability in older women: The Women's Health and Aging Study. *J Clin Epidemiol*. 1999;52(1):27-37.
- Jørgensen TL, Hallas J, Friis S, et al. Comorbidity in elderly cancer patients in relation to overall and cancer-specific mortality. *Br J Cancer*. 2012; 106(7):1353-1360.
- van Spronsen DJ, Janssen-Heijnen ML, Breed WP, et al. Prevalence of co-morbidity and its relationship to treatment among unselected patients with Hodgkin's disease and non-Hodgkin's lymphoma, 1993-1996. *Ann Hematol*. 1999;78(7):315-319.
- Sorror ML, Maris MB, Storb R, et al. Hematopoietic cell transplantation (HCT)-specific comorbidity index: A new tool for risk assessment before allogeneic HCT. *Blood*. 2005;106(8): 2912-2919.
- Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. *J Chronic Dis*. 1987;40(5):373-383.
- Sorror ML, Giralt S, Sandmaier BM, et al. Hematopoietic cell transplantation specific comorbidity index as an outcome predictor for patients with acute myeloid leukemia in first remission: Combined FHCRC and MDACC experiences. *Blood*. 2007; 110(13):4606-4613.
- Maruyama D, Fukuda T, Kato R, et al. Comparable antileukemia/lymphoma effects in nonremission patients undergoing allogeneic hematopoietic cell transplantation with a conventional cytoreductive or reduced-intensity regimen. *Biol Blood Marrow Transplant*. 2007; 13(8):932-941.
- Lim ZY, Ingram W, Brand R, et al. Impact of pretransplant comorbidities on alemtuzumab-based reduced-intensity conditioning allogeneic hematopoietic SCT for patients with high-risk myelodysplastic syndrome and AML. *Bone Marrow Transplant*. 2010;45(4):633-639.
- Fujimaki K, Sakai R, Fujisawa S, et al. Usefulness of hematopoietic cell transplantation-specific comorbidity index after allogeneic hematopoietic stem cell transplantation [Japanese]. *Gan To Kagaku Ryoho*. 2008;35(1):87-91.
- Majhail NS, Brunstein CG, Tomblyn M, et al. Reduced-intensity allogeneic transplant in patients older than 55 years: Unrelated umbilical cord blood is safe and effective for patients without a matched related donor. *Biol Blood Marrow Transplant*. 2008;14(3):282-289.

12. Boehm A, Sperr WR, Leitner G, et al. Comorbidity predicts survival in myelodysplastic syndromes or secondary acute myeloid leukaemia after allogeneic stem cell transplantation. *Eur J Clin Invest*. 2008;38(12):945-952.
13. Labonté L, Iqbal T, Zaidi MA, et al. Utility of comorbidity assessment in predicting transplantation-related toxicity following autologous hematopoietic stem cell transplantation for multiple myeloma. *Biol Blood Marrow Transplant*. 2008;14(9):1039-1044.
14. Pollack SM, Steinberg SM, Odom J, et al. Assessment of the hematopoietic cell transplantation comorbidity index in non-Hodgkin lymphoma patients receiving reduced-intensity allogeneic hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant*. 2009;15(2):223-230.
15. Kataoka K, Nannya Y, Ueda K, et al. Differential prognostic impact of pretransplant comorbidity on transplant outcomes by disease status and time from transplant: A single Japanese transplant centre study. *Bone Marrow Transplant*. 2010;45(3):513-520.
16. Barba P, Piñana JL, Martino R, et al. Comparison of two pretransplant predictive models and a flexible HCT-CI using different cut off points to determine low-, intermediate-, and high-risk groups: The flexible HCT-CI is the best predictor of NRM and OS in a population of patients undergoing allo-RIC. *Biol Blood Marrow Transplant*. 2010;16(3):413-420.
17. Pavlí J, Kew AK, Taylor-Roberts B, et al. Optimizing patient selection for myeloablative allogeneic hematopoietic cell transplantation in chronic myeloid leukemia in chronic phase. *Blood*. 2010;115(20):4018-4020.
18. Zipperer E, Pelz D, Nachtkamp K, et al. The hematopoietic stem cell transplantation comorbidity index is of prognostic relevance for patients with myelodysplastic syndrome. *Haematologica*. 2009;94(5):729-732.
19. Sperr WR, Wimazal F, Kundi M, et al. Comorbidity as prognostic variable in MDS: Comparative evaluation of the HCT-CI and CCI in a core dataset of 419 patients of the Austrian MDS Study Group. *Ann Oncol*. 2010;21(1):114-119.
20. Giles FJ, Borthakur G, Ravandi F, et al. The haematopoietic cell transplantation comorbidity index score is predictive of early death and survival in patients over 60 years of age receiving induction therapy for acute myeloid leukaemia. *Br J Haematol*. 2007;136(4):624-627.
21. Breccia M, Frustaci AM, Cannella L, et al. Comorbidities and FLT3-ITD abnormalities as independent prognostic indicators of survival in elderly acute myeloid leukaemia patients. *Hematol Oncol*. 2009;27(3):148-153.
22. Birninger N, Bornhäuser M, Schaich M, et al. The hematopoietic cell transplantation-specific comorbidity index fails to predict outcomes in high-risk AML patients undergoing allogeneic transplantation—investigation of potential limitations of the index. *Biol Blood Marrow Transplant*. 2011;17(12):1822-1832.
23. Sorror ML. Comorbidities and outcomes: Advancing the field comes at a price. *Biol Blood Marrow Transplant*. 2011;17(12):1721-1723.
24. Karnofsky DA, Burchenal JH. The clinical evaluation of chemotherapeutic agents in cancer. In: Macleod CM, ed. *Evaluation of Chemotherapeutic Agents*. New York: Columbia University Press; 1949:191-205.
25. Przepiorka D, Weisdorf D, Martin P, et al. 1994 Consensus conference on acute GVHD grading. *Bone Marrow Transplant*. 1995;15(6):825-828.
26. Cahn JY, Klein JP, Lee SJ, et al. Prospective evaluation of 2 acute graft-versus-host (GVHD) grading systems: A joint Société Française de Greffe de Moëlle et Thérapie Cellulaire (SFGM-TC), Dana Farber Cancer Institute (DFCI), and International Bone Marrow Transplant Registry (IBMTR) prospective study. *Blood*. 2005;106(4):1495-1500.
27. Carpenter PA. How I conduct a comprehensive chronic graft-versus-host disease assessment. [Commentary] *Blood*. 2011;118(10):2679-2687.
28. Dinakara P, Blumenthal WS, Johnston RF, et al. The effect of anemia on pulmonary diffusing capacity with derivation of a correction equation. *Am Rev Respir Dis*. 1970;102(6):965-969.
29. Cotes JE, Dabbs JM, Elwood PC, et al. Iron-deficiency anaemia: Its effect on transfer factor for the lung (diffusing capacity) and ventilation and cardiac frequency during sub-maximal exercise. *Clin Sci*. 1972;42(3):325-335.
30. Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas*. 1960;20(1):37-46.
31. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1971;76(5):378-382.
32. Streiner DL, Norman GR. *Health Measurement Scales: A Practical Guide to Their Development*. Oxford, United Kingdom: Oxford University Press; 2003.
33. Cohen J. Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70(4):213-220.
34. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas*. 1973;33(3):613-619.
35. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. *Psychol Bull*. 1969;72(5):323-327.
36. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
37. Hall SF, Groome PA, Streiner DL, et al. Interrater reliability of measurements of comorbid illness should be reported. *J Clin Epidemiol*. 2006;59(9):926-933.
38. Imamura K, McKinnon M, Middleton R, et al. Reliability of a comorbidity measure: The Index of Co-Existent Disease (ICED). *J Clin Epidemiol*. 1997;50(9):1011-1016.
39. Liu M, Domen K, Chino N. Comorbidity measures for stroke outcome research: A preliminary study. *Arch Phys Med Rehabil*. 1997;78(2):166-172.
40. Extermann M, Overcash J, Lyman GH, et al. Comorbidity and functional status are independent in older cancer patients. *J Clin Oncol*. 1998;16(4):1582-1587.
41. Tuijn S, Janssens F, Robben P, et al. Reducing interrater variability and improving health care: A meta-analytical review. [Review] *J Eval Clin Pract*. 2012;18(4):887-895.