NIH-PA Author Manuscript

A Revision of the Intensity of Treatment Rating Scale: Classifying the Intensity of Pediatric Cancer Treatment

Anne E. Kazak, Ph.D., ABPP[1,2], **Matthew C. Hocking, Ph.D.**[1], **Richard F. Ittenbach, Ph.D.**[3,4], **Anna T. Meadows, M.D.**[1,2], **Wendy Hobbie, MSN, PNP-BC, FAAN**[1,5], **Branlyn Werba DeRosa, Ph.D.**[1], **Ann Leahey, M.D.**[1], **Leslie Kersun, M.D.**[1,2], and **Anne Reilly, M.D.**[1,2]

[1]Division of Oncology, The Children's Hospital of Philadelphia

[2]Department of Pediatrics, University of Pennsylvania School of Medicine

[3]Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center

[4]Department of Pediatrics, the University of Cincinnati College of Medicine

[5]University of Pennsylvania School of Nursing

Abstract

BACKGROUND—We previously developed a reliable and valid method for classifying the intensity of pediatric cancer treatment. The Intensity of Treatment Rating Scale (ITR-2.0) [1] classifies treatments into four operationally defined levels of intensity and is completed by pediatric oncology specialists based on diagnosis, stage, and treatment data from the medical record. Experience with the ITR-2.0 and recent changes in treatment protocols indicated the need for a minor revision and revalidation.

METHODS—Five criterion raters reviewed the prior items, independently proposing additions and/or changes in the classification of diseases/treatments. Subsequent to a group discussion of the proposed changes, a revised 43 item ITR was evaluated. Pediatric oncologists ($n = 47$) completed a two-part online questionnaire. Validity of the classifications was determined by the oncologists classifying each disease/treatment into one of the four levels of intensity. Inter-rater reliability was calculated by having each oncologist classify the treatments of 12 sample patients using the new version which we call the ITR-3.

RESULTS—Agreement between median ratings of the 43 items for the pediatric oncologists and the criterion raters was high ($r = 0.88$). The median of the raters was either identical (81%) with the criterion ratings or discrepant by one level. Inter-rater reliability was very high when using the ITR-3 to classify 12 sample patients, with a median agreement of 0.90 and an intraclass correlation coefficient ($r_{ICC} = 0.86$).

CONCLUSIONS—With these minor modifications and updates, the ITR-3 remains a reliable and valid method for classifying pediatric oncology treatment protocols.

Keywords

pediatric oncology; measurement; treatment intensity; rating

Correspondence about this article, including requests for permission to use the ITR and information about its use, should be sent to Anne E. Kazak, Ph.D., ABPP, Division of Oncology, The Children's Hospital of Philadelphia, Room 1486 CHOP North, 34th St and Civic Center Blvd., Philadelphia, PA. 19104. 215-590-2220 (v), 215-590-3020 (f). (kazak@email.chop.edu).

INTRODUCTION

Current pediatric cancer treatments vary widely in their intensity, and are dependent upon the disease, stage, risk group and whether the disease is an initial diagnosis or relapse. While most clinical outcome studies in pediatric cancer are highly specific to a disease and treatment protocol, research on psychosocial and quality of life outcomes often necessitates the inclusion of patients with a wider range of diagnoses and treatments. The ability to classify the intensity of cancer treatments allows for comparisons across diagnostic groups.

Treatment intensity may be measured subjectively, from the perspective of the patient or family, or more objectively by healthcare providers based on medical data. From the family's perspective, treatment intensity is important, given its potential impact on quality of life and the related demands on the family. However, patient-reported appraisals of intensity can be highly individual and do not necessarily correspond to the more objective perspective of oncology specialists. In response to the need for a psychometrically reliable and valid means of classifying pediatric cancer treatment, we developed the Intensity of Treatment Rating Scale (ITR-2.0) [1]. Oncologists reviewed data abstracted from the medical record, Children's Oncology Group (COG) treatment protocol, disease stage and treatment modalities, and rated a patient's treatment using a four point scale, from least intensive to most intensive. There are specific criteria for each of the four levels. The classifications were validated by criterion raters at several pediatric oncology programs across the United States and showed very high levels of agreement (Median $r = 0.95$). Inter-rater reliability for the initial scale was 0.87. Inter-rater reliability in subsequent independent studies has been very strong ($r = 0.89$ to 0.96) across independent studies [2–6].

Treatment intensity ratings need updating periodically due to changes in treatment approaches and protocols [1]. Based on recent experience with the ITR-2.0 and observations that some diseases and treatments may not be optimally classified, we present a revision that reflects more contemporary treatments, using methods that paralleled the original scale development[1].

METHODS

Scale Revision

The Intensity of Treatment Rating scale (ITR-2.0) [1] is used to categorize the intensity of pediatric cancer treatment from least through most intensive based on treatment modality and stage/risk level for the patient. The ITR consists of two components: Intensity Levels and Content Items. Intensity Levels refers to the four categories of treatment intensity, from Level 1(minimally intensive) to Level 4 (most intensive). Content Items consisted of 34 different disease and/or treatment modalities, with each modality classified according to one of the four intensity levels. For example, Level 4 was most intensive and included treatments such as bone marrow transplantation or chemotherapy for acute myeloid leukemia.

In order to update the ITR-2.0 items and their classification to reflect current pediatric cancer protocols, a detailed review of its items was completed by clinical experts (criterion raters). A pediatric oncologist (AR) reviewed the existing classification of diseases and treatments and proposed changes to clarify, organize, and determine applicability to current cancer therapies. These changes were independently reviewed by two other pediatric oncologists (LK, AL). The first pediatric oncologist (AR) reviewed and accepted or clarified these changes. A fourth pediatric oncologist (ATM) and a pediatric oncology nurse practitioner (WH) then provided additional clinical expert review on this version. A meeting

of all five criterion reviewers was then convened to discuss nuances of the proposed changes and to finalize the revision.

Throughout this process, 23 items remained identical, 11 were added, 2 were removed, and changes were made to 9. Of the 11 new items, 2 were intended to capture low occurrence tumors not otherwise represented (e.g, "Tumor, other – 2 or 3 treatment modalities"). The 11 new items included 1 item in Level 1 (LCH, surgery or steroid injection only), 3 items in Level 2 (Langerhans Cell Histiocytosis [LCH] with chemotherapy; Thyroid cancer; Tumor, other – either chemo or radiation alone) 5 items in Level 3 (Biphenotypic leukemia – treated like ALL; Carcinoma NOS – two or more treatment modalities; Hemophagocytic lymphohistiocytosis [HLH], chemo alone; Soft tissue sarcoma – two or more treatment modalities; Wilms tumor [stages 3,4] – three treatment modalities; Tumor, other – 2 or 3 treatment modalities, and 2 items in Level 4 (Biphenotypic leukemia – treated like AML; Brain tumor – with HSCT). Modifications for clarity and specificity were made to 9 items. One item was changed to a lower level of intensity (Chronic Myeloid Leukemia – Chemotherapy Only). The criterion raters agreed that this set of 43 items was inclusive of most pediatric oncology diagnoses (Supplemental Table I).

Scale Validation

To validate the classification of treatments into four levels of intensity, pediatric oncologists were chosen at random from CHOP ($n = 23$) and from a roster of oncologists at other U.S. pediatric oncology centers ($n = 24$). Oncologists were recruited by email and asked to assist with a validation study of a new version of the ITR. The final sample represented 22 hospitals in 17 states across the United States. [1] Data from all 47 oncologists were used for analysis.[2]

This study was granted an exemption from the Committees for the Protection of Human Subjects at CHOP. Participants completed an online survey which first defined treatment intensity as "an over-arching evaluation that takes into account professionals' perceptions of the duration of therapy, side effects profile, risk of complications, number of agents, and treatment modalities, predicted time in the hospital, and the extent to which treatments are outpatient." The 43 disease/treatment items were presented in random order. The raters were instructed to rate independently each item by placing it into one of four levels of intensity:

---

[1]Oncologists were affiliated with 22 pediatric cancer programs across the United States that reflected variation in size and region.
Children's Hospital and Regional Medical Center (Seattle, WA)
Children's Healthcare of Atlanta (Atlanta, GA)
Children's Hospital (Denver, CO)
Children's Hospital (Oakland, CA)
Children's Hospital of Philadelphia (Philadelphia, PA)
Children's Memorial Medical Center (Chicago, IL)
Cincinnati Children's Hospital Medical Center (Cincinnati, OH)
City of Hope National Medical Center (Duarte, CA)
Dana-Farber Cancer Institute and Children's Hosp (Boston, MA)
New York University School of Medicine (New York NY)
Rainbow Babies and Childrens Hospital (Cleveland, OH)
Rhode Island Hospital (Providence, RI)
Tufts University (Boston, MA)
Tulane University Hospital and Clinic (New Orleans, LA)
University of Alabama (Birmingham, AL)
University of Arkansas (Little Rock, AR)
University of Minnesota Cancer Center (Minneapolis, MN)
University of Nebraska Medical Center (Omaha, NE)
University of Texas Southwestern Medical Center (Dallas, TX)
Vanderbilt University (Nashville, TN)
Washington University Medical Center (St Louis, MO)
Yale University School of Medicine (New Haven, CT)
[2]Data was missing from one oncologist for the second part of the questionnaire (rating the 12 case examples).

Level 1 (least intensive), Level 2 (moderately intensive), Level 3 (very intensive), or Level 4 (most intensive).

Inter-rater Reliability

The final step in this validation study was to determine the inter-rater reliability of the ITR when used to classify a patient's treatment into one of four categories. Data from 12 sample patients, both on and off treatment, were selected to represent several (three) treatments at each of the four categories equally. At least one patient description at each level represented a typical or common presentation while a second was considered to be "atypical." For example, in the moderately intensive level (Level 2), one patient example was included as a more common presentation or treatment (Yolk Sac Tumor, a standard regimen of cisplatin, etoposide, and bleomycin), as well as another judged to be less common (Anaplastic Large Cell Non-Hodgkin Lymphoma, treated with chemotherapy consisting of cyclophosphamide, methotrexate, and prednisone). The reason for doing so was to reflect the variability that would typically be seen in the patient population. Using diagnosis, stage/type, chemotherapy doses, and whether or not the patient received radiation, each oncologist used the ITR to classify the treatment intensity for each of 12 patient examples.

Data Analytic Plan

Spearman-rho correlation coefficients were computed to estimate inter-rater agreement between the 47 oncology raters and the criterion ratings. Median ratings across raters were computed and then correlated with the criterion ratings to get an overall measure of association across all items collectively. Individual ratings were correlated with criterion ratings for each rater and then summarized to assess the range of agreement between the sets of oncologist raters and the criterion raters. Empirically derived confidence intervals were generated when group medians differed from criterion ratings, based on order statistics, normal distribution theory, and correction for continuity [7].

To evaluate how the updated ITR performed using actual case examples, inter-rater reliability estimates were generated for 12 patient cases using the same pool of raters. Two measures of association were obtained: Kendall's Tau-b was used to estimate the relationship between each rater's assessment and the set of 12 case ratings, and a two-way randomized, intraclass correlation coefficient estimated consistency across raters for the same set of 12 clinical cases. The intraclass coefficient (ICC) was estimated using the more stringent assumption of absolute agreement across all 46 raters due to the need for raters to evaluate the case examples in highly comparable ways as well as to account for any systematic differences in raters that might be present. Power and sample size estimates were computed at the design phase of the study and required a minimum of 42 respondents for the correlation phase of the study and 46 respondents for the ICC phase of the study. These calculations were based on an $r = 0.90$ (for the Spearman r, LL[lower limit] $= 0.80$), $r = 0.80$ (for the r_{ICC}, ω [distance limit] $= 0.19$), and an $\alpha = 0.99$ for both analyses. Because the two phases of the study required different sample sizes to detect statistical significance, the sampling frame was predicated on the larger, more conservative sample size of 46 respondents. All analyses were conducted using SPSS v18 (IBM Corp., Somers, NY) and STATA v10.0 (STATACorp, College Station, TX) software.

RESULTS

Scale Validation

The agreement between the median ratings of the oncologists and the criterion ratings was high for the 43 items on the ITR-3 ($r = 0.88$, range 0.66 to 0.88). The median and the criterion ratings were identical for 35 of the 43 items (81%). For each of the eight items that

was not identical, the median rating was one level below the criterion rating indicating that the raters considered the treatments associated with those diseases as less intense than the criterion rating. Additionally, five of the eight items that were not identical fell within the empirically derived 95% confidence interval. The criterion rating fell outside the empirically derived confidence interval for the following three diseases: Hemophagocytic lymphohistiocytosis (HLH), chemo alone; Wilms Tumor (Stages 3, 4); and Relapsed Disease - Excluding Hodgkin Lymphoma or first relapse of Wilms Tumor. Of these items, only Hemophagocytic lymphohistiocytosis (HLH) represents a new item that was not in the previous version of the ITR.

The criterion raters met to review each of the eight items where the criterion rating differed from that of the raters. For the five items that were within the 95% derived confidence interval, it was decided to retain the criterion rating but to modify two of the items slightly for clarity. Specifically, *Carcinoma NOS* was changed to "Carcinoma NOS with two or more treatment modalities" and *Soft Tissue Sarcoma unless surgery alone* was changed to "Soft Tissue Sarcoma with two or more treatment modalities." For the three ratings outside the confidence interval, the majority of participants (69, 71, 76%, respectively) picked the median rating for the item. Therefore, the criterion rating was in the clear minority and the classification of the item on the ITR-3 was changed to reflect the data. One additional change made was to change *Wilms Tumor Stages 3,4 to* Wilms Tumor Stages 3, 4 with three treatment modalities.

Inter-rater reliability

Examining inter-rater reliability of the oncologists' ratings of the 12 patient examples using Kendall's Tau-b revealed a high median level of agreement between the criterion rating and each oncologist ($r = 0.90$, range 0.66 to 1.0,). A second measure of inter-rater reliability based on the intraclass correlation coefficient also indicated a high level of agreement and reliability across all 46 raters, collectively ($r_{ICC} = 0.86$).

DISCUSSION

The Intensity of Treatment Rating Scale remains a psychometrically strong measure of intensity of treatment for pediatric oncology treatments. The revised ITR-3 shows psychometric properties consistent with the prior version. Even with a larger pool of raters in this study, oncologists classified the majority of treatments consistently with the criterion raters or within the confidence interval of the criterion ratings. The inter-rater reliabilities are very strong, again accounting for a larger sample of raters than in the prior version. In summary, we recommend this version (ITR-3) of the scale over the previous one. As with the earlier version, accuracy of retrieval of treatment information is essential in using the ITR-3. In our experience this step should be completed by an oncologist or nurse practitioner familiar with potential nuances in medical record data.

Prior attempts to classify treatment intensity are variable and generally dependent on the nature of the study. Measures of treatment intensity are incorporated in investigations specific to only one disease group [8] or to particular types of treatments, such as treatments directed at the central nervous system [9]. Reports from the Childhood Cancer Survivor Study (CCSS) utilize detailed treatment information and individual treatment predictors (e.g., platinum compound dose) on outcomes of interest, such as auditory [10] or gastrointestinal complications [11], but do not use an overall rating of treatment intensity. A series of studies [12–14] have used a forced-choice technique where oncologists rate protocols from least to most intense, however this procedure has not been as thoroughly evaluated empirically as the ITR-3 and is more prone to subjectivity. The ITR-3 fills this

void in pediatric oncology research by providing a broad, validated measure of treatment intensity that can be used in studies with a variety of pediatric cancer diagnoses.

The previous version of the ITR has been used to characterize samples in terms of the cancer treatment received [2], to establish the comparability of groups in psychosocial intervention studies [4,15] and to compare a subscale with a full sample [16]. The ITR has also been used to assess the extent to which the treatment intensity relates to psychosocial and medical outcomes. For example, treatment intensity is positively related to survivor perceptions of late effects [17] and to psychological outcomes and beliefs about health in adolescent and young adult survivors [3]. Alternatively, in some studies treatment intensity as measured by the ITR-2.0 is not associated with psychosocial risk in families at diagnosis [5,15] or for risk for depression in adolescents with cancer[18]. Treatment intensity also did not contribute significantly to quality of life in adolescents with cancer [19]. These "negative" findings may also be important in further understanding how treatment related variables may contribute to psychosocial and quality of life outcomes.

Given consistent progress in clinical trials, modifications to treatment protocols, and advances in treatment-related side effects in children's cancer, evaluation of the scale will be necessary at regular intervals in order to assure that the ratings reflect current therapies. The process of revising the ITR was an iterative process that highlighted some of the complexities and potential discrepancies, in how experts view the diseases and treatments, including some variability in the raters' evaluations of the treatments. The ITR-3 is based on current treatment protocols in the United States. Although COG protocols are used widely around the world, there are some differences among standard treatments by country or region that warrant some caution in using the ITR-3 outside of North America. In conclusion, the ITR-3 is an easily used reliable and valid method for used rating for classifying the intensity of current treatment protocols in pediatric oncology.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Werba BE, Hobbie WL, Kazak AE, et al. Classifying the intensity of pediatric cancer treatment protocols: The Intensity of Treatment Rating Scale 2.0 (ITR-2). Pediatric Blood and Cancer. 2007; 48:673–677. [PubMed: 17427232]

2. Alderfer MA, Mougianis I, Barakat LP, et al. Family psychosocial risk, distress, and service utilization in pediatric cancer: Predictive validity of the Psychosocial Assessment Tool. Cancer. 2009; 115 (18 suppl):4339–4349. [PubMed: 19731359]

3. Kazak AE, DeRosa B, Schwartz L, et al. Psychological outcomes and health beliefs in adolescent and young adult (AYA) survivors of childhood cancer and controls. Journal of Clinical Oncology. 2010; 28:2002–2007. [PubMed: 20231679]

4. Lutz Stehl M, Kazak AE, Alderfer MA, et al. Conducting a randomized clinical trial of an psychological intervention for parents/caregivers of children with cancer shortly after diagnosis. Journal of Pediatric Psychology. 2009; 34:803–816. [PubMed: 19091806]

5. Pai ALH, Patino-Fernandez AM, McSherry M, et al. The Psychosocial Assessment Tool (PAT2.0): Psychometric properties of a screener for psychosocial distress in families of children newly diagnosed with cancer. Journal of Pediatric Psychology. 2008; 33:50–62. [PubMed: 17609228]

6. McCarthy MC, Clarke NE, Vance A, et al. Measuring psychosocial risk in families caring for a child with cancer: The psychosocial assessment tool (PAT2.0). Pediatric Blood and Cancer. 2009; 53:78–85. [PubMed: 19343796]

7. Gibbons, JD. Nonparametric methods for quantitative analysis. 3. Columbus, OH: American Sciences Press; 1985.

8. Chow EJ, Pihoker C, Hunt K, et al. Obesity and hypertension among children after treatment for acute lymphoblastic leukemia. Cancer. 2007; 110:2313–2320. [PubMed: 17896787]

9. Vannatta K, Gerhardt CA, Wells RJ, et al. Intensity of CNS treatment for pediatric cancer: Prediction of social outcomes in survivors. Pediatric Blood and Cancer. 2007; 49:716–722. [PubMed: 17096410]

10. Whelan K, Stratton K, Kawashima T, et al. Auditory complications in childhood cancer survivors: A report from the Childhood Cancer Survivor Study. Pediatric Blood and Cancer. 201110.1002/pbc.23025

11. Goldsby R, Chen Y, Raber S, et al. Survivors of childhood cancer have increased risk for gastrointestinal complications later in life. Gastroenterology. 201110.1053/j.gastro.2011.01.049

12. Gerhardt CA, Vannatta K, Valerius KS, et al. Social and romantic outcomes in emerging adulthood among survivors of childhood cancer. Journal of Adolescent Health. 2007; 40:462, e469–462, e415. [PubMed: 17448406]

13. Noll RB, Garstein MA, Vannatta K, et al. Social, emotional, and behavioral functioning of children with cancer. Pediatrics. 1999; 103:71–78. [PubMed: 9917442]

14. Thompson AL, Gerhardt CA, Miller KS, et al. Survivors of childhood cancer and comparison peers: The influence of peer factors on later externalizing behavior in emerging adulthood. Journal of Pediatric Psychology. 2009; 34:1119–1128. [PubMed: 19324936]

15. Kazak AE, Barakat LP, DiTaranto S, et al. Screening for psychosocial risk at cancer diagnosis: The Psychosocial Assessment Tool (PAT). Journal of Pediatric Hematology and Oncology. 2011; 33:289–294.

16. Lutz Stehl M, Kazak AE, Hwang W-T, et al. Innate immune markers in mothers and fathers of children newly diagnosed with cancer. Neuroimmunomodulation. 2008; 15:102–107. [PubMed: 18679048]

17. Schwartz L, Mao J, Werba BE, et al. Self-reported health problems of young adults in clinical settings: Survivors of childhood cancer and healthy controls. Journal of the American Board of Family Medicine. 2010; 23:306–314. [PubMed: 20453176]

18. Kersun L, Rourke MT, Mickley M, et al. Screening for depression and anxiety in adolescent cancer patients. Journal of Pediatric Hematology and Oncology. 2009; 31:835–839.

19. Barakat LP, Marmer P, Schwartz LA. Quality of life among adolescents with cancer: Family risks and resources. Health and Quality of Life Outcomes. 2010; 8:63. [PubMed: 20584303]