MEASUREMENT INVARIANCE IN HEALTH-RELATED QUALITY OF LIFE ASSESSMENT

Analysis Concept Proposal

August 16, 2011

1. STUDY TITLE

Measurement invariance in health-related quality of life between young adult survivors of childhood cancer and their siblings

2. WORKING GROUP AND INVESTIGATORS

This proposed publication will be within the Psychology Committee. Proposed investigators will include:

I-Chan Huang	UFCOM	ichuang@ufl.edu	(352) 265-2514
Wendy Leisenring	FHCRC	wleisenr@fhcrc.org	(206) 667-4374
Gwendolyn Quinn	MCC	gwen.quinn@moffitt.org	(813) 745-1359
Elizabeth Shenkman	UFCOM	eas@ichp.ufl.edu	(352) 265-2547
Kelly Kenzik	UFCOM	kkenzik@ichp.ufl.edu	(352) 265-2572
Mandy Li	BC	zhushan.li@bc.edu	(617) 552-4534
Tara Brinkman	SJCRH	tara.brinkman@stjude.org	(901) 595-5683
Christopher Recklitis	DFCI	christopher_recklitis@dfci.harvard.edu	(617) 632-3839
Greg Armstrong	SJCRH	greg.armstrong@stjude.org	(901) 595-5892
Les Robison	SJCRH	les.robison@stjude.org	(901) 595-6078
Kevin Krull	SJCRH	kevin.krull@stjude.org	(901) 595-5891

Note:
UFCOM: University of Florida College of Medicine
SJCRH: St Jude Children's Research Hospital
BC: Boston College
MCC: Moffitt Cancer Center
DFCI: Dana-Farber Cancer Institute
FHCRC: Fred Hutchinson Cancer Research Center

3. BACKGROUND AND RATIONALE

For the last three decades, the 5-year survival rate for childhood cancer has improved substantially in the United States, from less than 50% in the 1970s to 80% today.[1] An estimated 11.7 million Americans are cancer survivors,[2] and more than 330,000 are survivors of childhood cancer with 24% of them having survived greater than 30 years.[3] Childhood cancer survivors are at great risk of developing late effects (LEs), which impair health-related quality of life (HRQOL) of young adult survivors of childhood cancer (YASCC).[4]

To understand the state of HRQOL impairment among YASCC, it is important to identify appropriate control groups and compare HRQOL of YASCC to the control groups. In previous YASCC studies, age- and sex-matched siblings and/ or representative populations (i.e., the norms) were chosen to serve as control groups. However, evidence is mixed with regard to HRQOL outcomes between YASCC and control groups.

- In the VOLG project, Stam and colleagues compared HRQOL (SF-36) of 353 Dutch YASCC to 507 peers and reported that HRQOL between the two groups was comparable. Specifically, the effect size in physical component scores (PCS) and mental component scores (MCS) of the SF-36 were 0.15 and 0.12, respectively.[5]

- In the Canadian Childhood Cancer Surveillance and Control Program (CCCSCP), Maunsell and colleagues demonstrated that 1,334 YASCC had a small to moderate impairment in HRQOL (SF-36) compared to 1,477 age- and sex-matched general populations. The impairment on these two groups was more salient on physical aspects of HRQOL (e.g., physical functioning, role limitation due to physical health problem, and general health) than mental aspects (e.g., bodily pain, vitality, and role limitation due to mental health problem).[6]
- In the British Childhood Cancer Survivor Study (BCCSS), Reulen and colleagues demonstrated that YASCC had comparable SF-36 MCS compared to UK-norms (effect size = 0.1). By contrast, the difference on the SF-36 PCS varied by age, where YASCC aged between 16 and 19 years old scored similarly to the UK-norms and YASCC aged 25 years old and above scored statistically and clinically significantly below the UK-norms. Survivors of central nervous system (CNS) and bone tumors scored significantly below on the SF-36 PCS than the UK-norms.[7]
- In the Childhood Cancer Survivor Study (CCSS), Zeltzer and colleagues demonstrated that comparing 7,147 YASCC to 388 age- and sex-matched siblings and the US norms, YASCC had greater impaired HRQOL in the SF-36 physical functioning, role limitation due to physical health problem, general health, and social functioning compared to siblings. Similarly, YASCC had greater impairment in the domains of physical functioning, role limitation due to physical health problem, general health, vitality, role limitation due to emotional health problem, and social functioning, but slightly better in mental health compared to the US norms. Survivors of CNS tumors, lymphoma, bone and soft tissue sarcoma gave the lowest HRQOL scores in physical domains.[8]

The factors related to more positive HRQOL outcomes among YASCC compared to the control groups are complex and largely unknown. We generally believe that YASCC's psychosocial adjustment related to their past and current cancer experience as well as their expectation and outlook for the future might play a critical role. The psychosocial adjustment potentially leads YASCC to raise their HRQOL ratings.

On the one hand, these findings may reflect the occurrence of psychological resilience or posttraumatic growth developed by YASCC in coping with the adverse circumstances.[9] Many YASCC report benefits from their cancer experience ranging from an increased appreciation each day to greater feelings of personal growth such as satisfaction with global well-being compared to healthy comparison groups.[10, 11] Qualitative studies suggest that 20% of YASCC report feeling calm and mentally stronger, and having good self-confidence. More than 25% of YASCCs report a fuller appreciation of life, having a more positive view of life, and making different priorities.[12] On the other hand, these findings may suggest the presence of repressive adaptation where YASCC tend to deny difficulties and enhance self-appraisal, leading them to report more positive health and overestimate their subjective well-being and satisfaction with daily life.[10, 13, 14] Several studies using a life course approach demonstrated that YASCC are more likely to reach fewer milestones, such as leaving parent's homes, having a first boyfriend/ girlfriend, or achieving these milestones at an older age than their peers.[15] In this context, YASCC may respond to a HRQOL survey in a socially desirable manner. Social desirability may be affecting the responses toward overestimation of HRQOL ratings because subjects may feel it is important to appear function similar to their peers.[14, 16]

Indeed, perception of HRQOL among cancer survivors may change over time. From a longitudinal point of view, the benefits of psychosocial adjustment may decline as time since

diagnosis increases and late effects appear. With advancing age, there is increased risk of additional major life events, functional limitations or chronic health conditions, which may lead to impaired HRQOL.[17] In a CCSS study, Zeltzer and colleagues reported that YASCC who were 25 years from diagnosis had better HRQOL in different domains than those being treated 30 years from diagnosis.[8, 18] Another CCSS study conducted by Zebrack and colleagues suggests that YASCC were significantly more likely than siblings to report a perceived positive impact. This finding is especially salient among survivors who had fewer years since diagnosis.[19]

If the research goal is to compare HRQOL between YASCC and control groups, from measurement perspective, it is important to demonstrate whether items of the HRQOL instrument operate equivalently between YASCC and control groups (i.e., **measurement invariance**). Given the potential phenomenon of psychological resilience, posttraumatic growth, and/ or repressive adaptation experienced by YASCC, it is possible that YASCC rate or interpret the items in different ways compared to control groups. Without demonstrating measurement invariance, HRQOL assessment between YASCC and control groups may be misleading because we cannot rule out whether the disparity in HRQOL among the two groups reflects an unbiased measurement or not. In addition, it is unclear whether YASCC who perceive their cancer experience as more beneficial in personal growth will lead them to respond to HRQOL items in more positive direction compared to YASCC who perceive their cancer experience as less beneficial given the same level of underlying health conditions. Similarly, it is unclear whether YASCC of short-term survivors will respond to HRQOL items in a more positive direction compared to YASCC of long-term survivors given the same level of underlying health conditions.

Differential item functioning (DIF) analysis is an item-level psychometric method to investigate measurement invariance between different study groups by exploring whether the likelihood of responding to an item between different groups is the same or not, while conditioning on the same level of the underlying HRQOL.[20, 21] Theoretically, if the underlying HRQOL is the same between a YASCC and a sibling, one should expect both subjects will have the same probability of responding to a particular category on an item (e.g., “never” have a problem with walking more than one block). A DIF exists when this assumption is not held. DIF may lead to overestimating or underestimating the HRQOL score of a YASCC, thus mistakenly classifying a YASCC to different levels of health status.

Although the psychometric properties of the SF-36 have been evaluated in our and other studies,[22-24] these studies are based on YASCC population alone without accounting for the variation of control groups. Several psychometric methods have been developed for DIF analysis. Teresi classified these methods by nonparametric and parametric methods.[20, 21] Parametric methods, which are frequently used in HRQOL research, include item response theory-likelihood ratio (IRT-LR) method, ordinal logistic regression (OLR) method, **multiple indicator-multiple cause (MIMIC) method**, and differential functioning of items and tests (DFIT). The MIMIC method receives more attention in recent studies because this method can model item response function and group difference in underlying HRQOL simultaneously.[25, 26] Importantly, the MIMIC method can accommodate the background variables (confounding variables) into DIF analysis, which allows a meaningful comparison of latent HRQOL scores among different groups.

The main purpose of this study is to test DIF in each domain of the SF-36 between YASCC and their siblings. Uniform and non-uniform DIF will be specifically identified based on a **multiple group-MIMIC (MG-MIMIC)** methodology (see Analytic Approach section). In

this study, we define young adults as those who were aged between 18 and 40 years of age at the time of survey completion. Data collected from CCSS 2003 Long-Term Follow-Up Study (2003 LTFU) will be used for analysis. In addition, we will assess DIF between short-term survivors (less than 25 years) and long-terms survivors (more than 25 years) in each domain of the SF-36. The impact of DIF will be evaluated by examining the expected item/ test scores and item/ test information functioning between both groups. Finally, we will examine the change in domain scores of each subject before and after accounting for DIF items in the score calculation (i.e., **DIF calibration**; see Analytic Approach section), and test the discrepancy in domain scores between both groups before and after DIF calibration.

4. SPECIFIC AIMS/OBJECTIVES/RESEARCH HYPOTHESES

- Aim 1 (primary aim): to test DIF between YASCC and siblings on each domain of the SF-36 given the same level of underlying HRQOL and background variables.

  Hypothesis: YASCC tend to demonstrate more positive response in the DIF items compared to siblings given the same level of underlying HRQOL and background variables. This DIF finding may reflect the phenomenon of psychological resilience, posttraumatic growth, or repressive adaptation experienced by YASCC. In addition, more DIF items will be observed in mental domains of the SF-36 (vitality, social functioning, role-limitation due to mental health problem, and mental health) compared to physical domains of the SF-36 (physical functioning, role-limitation due to physical health problem, pain, and general health). This is because the content of the mental aspects of HRQOL is more subjective than the physical aspects.

- Aim 2 (secondary aim): to test DIF between high and low perceived positive impact (or personal growth) on each domain of the SF-36 given the same level of underlying HRQOL and background variables (including severity of chronic conditions).

  Hypothesis: YASCC who perceive their cancer experience as more beneficial in personal growth tend to demonstrate more positive response in the DIF items compared to YASCC who perceive cancer experience as less beneficial given the same level of underlying HRQOL and background variables. This is may be due to the fact that YASCC with higher perceived positive impact may adjust the impact of the cancer experience on their daily functioning more so than YASCC with lower perceived positive impact.

- Aim 3 (exploratory aim): to test DIF between short-term (< 25 years) and long-term survivors (≥ 25 years) on each domain of the SF-36 given the same level of underlying HRQOL and background variables (including severity of chronic conditions).

  Hypothesis: We will test this specific aim based on our working hypothesis. We specifically hypothesize that short-term survivors tend to demonstrate more positive response in the DIF items compared to long-term survivors given the same level of underlying HRQOL and background variables. This is because the short-term survivors may adjust the psychosocial impact associated with cancer experience on daily functioning better than long-term survivors.

5. METHODS

1) Subjects:

a. YASCC: Survivors in the CCSS cohort who completed the long version of the 2003 Long-Term Follow-Up Study (2003 LTFU) with SF-36 instrument variables and were between 18 and 40 years of age at the time of survey completion.
b. Control group: Siblings of YASCC in the CCSS cohort who completed the long version of the 2003 Long-Term Follow-Up Study Sibling Survey (2003 LTFU Sibling) with SF-36 variables and were between 18 and 40 years of age at the time of survey completion.

2) Outcomes of interests:

Self-report HRQOL measured by the Medical Outcomes Study SF-36 instrument (the SF-36) (item #E1 through #E22 and #F1 through #F14 in both 2003 LTFU and 2003 LTFU Sibling). The SF-36 is comprised of 36 items measuring eight domains of HRQOL: physical functioning, role-limitation due to physical health problems, bodily pain, general health, vitality, social functioning, role-limitation due to emotional health problems, and general health. Because DIF is a type of item-level analysis, all items in the SF-36 will be used in this study.

3) Variables associated with DIF in HRQOL
   a. YASCC and sibling (Aim 1).
   b. High and low perceived positive impact (or personal growth) (Aim 2)
      i. Post-traumatic growth inventory (2003 LTFU; items H1-H21). Given the fact that no large representative norms are available, we will use data collected from sibling to generate the cutoffs. Specifically, we will calculate the raw scores for siblings and then linearly transform to T-scores with a mean 50 and a standard deviation 10. The T-score associate with the top 10 percentile will be used as a cutoff to define high and low perceived positive impact for YASCC.
   c. Long-term (≥ 25 years) and short-term (< 25 years) survivors (Aim 3).

4) Confounding variables (see Table 1 in Appendix):
Of note, we will examine the missingness of each variable by YASCC and siblings. If the missingness of a specific variable (e.g., incomes) is significant, we will not include this variable in the analysis.

   a. Socio demographic
      i. Age: 18 – 40 years old for both YASCC and siblings.
      ii. Gender: male and female.
      iii. Race/ethnicity: White, non-Hispanic; Black, non-Hispanic; Hispanic; and other.
      iv. Education: below high school; high school graduate/ GED; some college/ training after high school; college graduate; postgraduate level; and other (item #1 in both 2003 LTFU and 2003 LTFU Sibling).
      v. Marital status: married/ living with a partner; widowed/ divorced/ separated; and single (item #2 in both 2003 LTFU and 2003 LTFU Sibling).
      vi. Living arrangement: live with spouse/ partner; live with parents; live with roommate; live with brothers/ sisters; live with other relatives; live alone; and other (item #3 in both 2003 LTFU and 2003 LTFU Sibling).
      vii. Employment status: working full-time; working part-time; and other (item #4 in both 2003 LTFU and 2003 LTFU Sibling).

viii. Insurance status: insured; uninsured; and other (Canadian resident) (item #M1, M1a, and M1b in both 2003 LTFU and 2003 LTFU Sibling).
ix. Incomes: > $19,999; $20,000 – $59,999; $60,000 – $79,999; $80,000 – $99,999; and ≤ $100,000 (item #S1 through #S3 in both 2003 LTFU and 2003 LTFU Sibling).

b. Weight and height: underweight (BMI<18.5 kg/m^2); normal weight (BMI: 18.5 – 24.9 kg/m^2); overweight (BMI: 25.0 – 29.9 kg/m^2); and obese (BMI: ≥30 kg/m^2) (item #7 and #8 in both 2003 LTFU and 2003 LTFU Sibling).
c. Cancer diagnosis
    i. Primary cancer: leukemia; central nervous system (CNS) tumor; Hodgkin lymphoma; Non-Hodgkin lymphoma; Wilms tumor; neuroblastoma; soft tissue sarcoma; bone tumor; and other.
    ii. Second cancer: yes/ no (item #R1 in both 2003 LTFU and 2003 LTFU Sibling).
d. Survival time (will be generated using the following two variables)
    i. Age at diagnosis: in years.
    ii. Age at interview: in years.
e. Cancer treatment
    i. Chemotherapy: none; methotrexate; corticosteroid; anthracyclines; alkylating agents; and other chemotherapy (all yes/ no).
    ii. Radiotherapy: none; cranial radiotherapy; and other radiotherapy (all yes/ no).
    iii. Surgery: none; amputation; and other surgery (all yes/ no).
f. Health status and psychological outcomes
    i. Type and severity of chronic conditions:
        1. Type of chronic conditions: major joint replacement; congestive heart failure; second malignant neoplasm; cognitive dysfunction, severe; coronary artery disease; cerebrovascular accident; renal failure or dialysis; hearing loss not corrected by aid; legally blind or loss of an eye; ovarian failure; and other.
        2. Grading of chronic condition according to Oeffinger 2006 – mild (Grade 1); moderate (Grade 2); severe (Grade 3); or life-threatening or disabling (Grade 4). Each YASCC may possess a variety of chronic conditions with different grades. However, the high grade will be used to represent the severity of chronic conditions of a specific YASCC.
    ii. Psychological symptoms:
        1. Brief Symptom Inventory (BSI-18): for each YASCC, T-scores on three subscale scores (anxiety, depression, and somatization) and a summary scale (global severity index; GSI) will be generated. Each YASCC, the cutoff of 63 on each subscale and a summary scale will be used to dichotomize the level of symptoms[27] (item #G1 through #G20 in both 2003 LTFU and 2003 LTFU Sibling).
        2. Post-Traumatic Stress Disorder (PTSD) symptoms: for each YASCC, DSM IV diagnostic requirements of at least one re-experiencing symptom, two arousal symptoms, and three avoidance symptoms will be used to dichotomize the level of PTSD[28] (item #K1 through #K17 in both 2003 LTFU and 2003 LTFU Sibling).

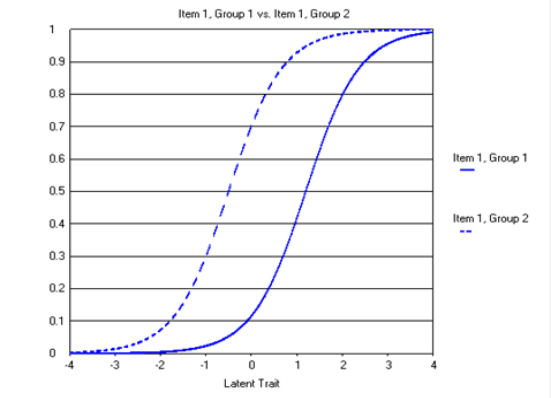
5) Analytic approach**:**

For Aim 1: DIF in HRQOL between YASCC and siblings

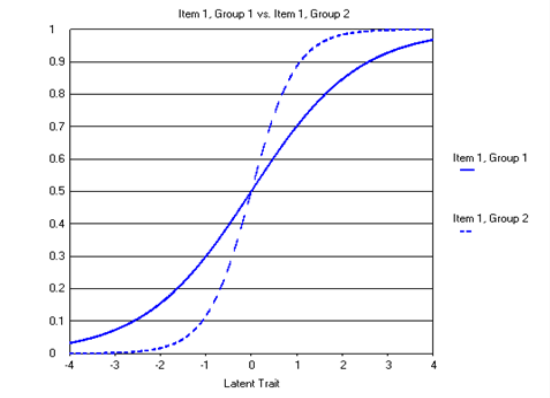
We will conduct dimensionality assessment for each domain of the SF-36 prior to DIF analysis. We hypothesize that the potential source of DIF is the different perceptions of HRQOL by YASCC and their siblings. It is likely that an item (e.g., walking more than one mile) designed to measure a specific functional status (i.e., physical functioning of the SF-36) is essentially relevant to multiple concepts of HRQOL (e.g., physical functioning, vitality, and general health of the SF-36). It, therefore, will violate the unidimensionality assumption of instrument design, and can be obvious among YASCC in part due to the perceived dependency across different domains and the adaptive style after the illness. We will assess the dimensionality of each domain in YASCC and sibling groups, respectively, using a standard confirmatory factor analysis (CFA).

DIF occurs when an item performs differently between the groups *given the same level of underlying HRQOL*. In this study, we will use a MG-MIMIC method to identify DIF associated with cancer survivorship. The MG-MIMIC is a special case of the MG-CFA methodology, where the MG-MIMIC is specifically allowed to control for the influence of background variables (confounding variables) on the DIF analysis. In this study, the following background variables will be included in MG-MIMIC model: age, gender, race/ethnicity, educational background, and severity of chronic conditions and psychological symptoms. Serial tests of nested models, beginning with the most constrained model, sequentially relaxing cross-group equality constraints on the parameters, and ending up with the least constrained model, are performed to detect uniform and non-uniform DIF. **Uniform DIF** is captured by the discrepancy in thresholds of a categorical item between both groups (e.g., YASCC and siblings) and **non-uniform DIF** is captured by the discrepancy in the loadings (or slopes) of an item on underlying HRQOL between both groups. Figure 1 provides an intuitive interpretation for DIF analysis using items with dichotomous response categories as an example. Group heterogeneity in HRQOL is indicated by the discrepancy on the mean of the underlying scores calculated by the model. The MG-MIMIC method uses the purification procedure to identify the anchor items in DIF assessment. Anchor items are the non-DIF items that are invariant in item parameters between both groups.

Figure 1: Intuitive interpretation of DIF analysis (dichotomous items as an example)



Uniform DIF



Non-uniform DIF

Note: X-axis: underlying score of a specific measure (e.g., physical functioning); Y-axis: probability of endorsing a specific category on an item.

Technically, the model building procedures for detecting DIF are the same as for a single group-CFA with covariates. The procedures of DIF analysis are iterative and inclusive of the following steps:

- Step 1: Estimate a baseline model which is fully invariant in factor loadings, thresholds, and residual variances of the items, variance of latent traits, and scaling factors. The only invariant parameter is the means of the latent variables between both groups which allow estimating the group differences in underlying HRQOL;

- Step 2: Examine the model modification indices (MIs) for the baseline model and identify the modification that would result in the largest improvement in model fit based on factor loadings and thresholds of items;

- Step 3: Use the DIFFTEST procedure[29] to fit a model that relaxes the constraint on factor loadings relative to the baseline model (i.e., non-uniform DIF identification);

- Step 4: Use the DIFFTEST procedure[29] to fit a model that relaxes the constraint identified in item thresholds relative to the baseline model (i.e., uniform DIF identification);

- Step 5: Compare the chi-square values from DIFTEST procedure for these two modifications to identify the largest one, and if it is significant, accept that modification and reject the other (note, a model in Step 5 becomes a new baseline model);

- Step 6: Estimate this new baseline model, examine the MIs, and repeat Steps 2 through 6 until there are no longer any significant model modifications were identified.

  Table 2 in Appendix shows the results of DIF identification and the parameters corresponding to DIF status of individual items.

In addition to the DIF tests, we will examine the magnitude of DIF visually by plotting the expected item score function (defined as a subject's expected response to an item across the underlying HRQOL continuum) and item information function (defined as measurement precision of an item across the underlying HRQOL continuum) between both groups (see Figure 2 in Appendix). We will also plot the expected test score function and test information function to investigate the magnitude of DIF at the aggregate (i.e., domain) level.

We will assess the impact of DIF on the change of domain scores for each subject before and after DIF calibration (see Table 3 in Appendix). We will examine whether the score change is above two points (equivalent to 0.2 unit of effect size) as the evidence of minimally important change. Further, we tested the discrepancy in the underlying domain scores between both groups, and compared the discrepancy before and after the DIF calibration. The criteria < 0.2, 0.2-0.49, 0.5-0.79, and > 0.8 will be used to indicate negligible, small, moderate, and large difference, respectively.[30]

<u>For Aims 2 and 3: DIF in HRQOL between high and low perceived positive impact and between short-term and long-term survivors</u>

We will use exactly the same methodology to conduct DIF analysis for Aims 2 and 3, by replacing the variable of interests from “YASCC versus siblings” (in Aim 1) to “high versus low perceived positive impact” using the cutoff (derived from sibling group) on the domain score of the Perceived Positive Impact measure (in Aim 2), and “short-term versus long-term” (cutoff: 25 years) survivors (in Aim 3).

6) Software**:**

For Aims 1, 2, and 3, we will perform dimensionality assessment and DIF analyses using Mplus 6.0, and conduct the rest of analyses using Stata 9.0. The analysis will be conducted by PI: I-Chan Huang at the University of Florida College of Medicine, with review of results and manuscript carried out by members of the CCSS Statistical Center.

6. REFERENCES

1. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2009. CA Cancer J Clin 2009;59:225-249.
2. Altekruse SF, Kosary CL, Krapcho M, et al. SEER Cancer Statistics Review, 1975–2007. Bethesda, MD: National Cancer Institute; 2009.
3. Mariotto AB, Rowland JH, Yabroff KR, et al. Long-term survivors of childhood cancers in the United States. Cancer Epidemiol Biomarkers Prev 2009;18:1033-1040.
4. National Cancer Policy Board (U.S.), Weiner SL, Simone JV. Childhood cancer survivorship: improving care and quality of life. Washington, D.C.: National Academies Press; 2003.
5. Stam H, Grootenhuis MA, Caron HN, et al. Quality of life and current coping in young adult survivors of childhood cancer: positive expectations about the further course of the disease were correlated with better quality of life. Psychooncology 2006;15:31-43.
6. Maunsell E, Pogany L, Barrera M, et al. Quality of life among long-term adolescent and adult survivors of childhood cancer. J Clin Oncol 2006;24:2527-2535.
7. Reulen RC, Winter DL, Lancashire ER, et al. Health-status of adult survivors of childhood cancer: a large-scale population-based study from the British Childhood Cancer Survivor Study. Int J Cancer 2007;121:633-640.
8. Zeltzer LK, Lu Q, Leisenring W, et al. Psychosocial outcomes and health-related quality of life in adult childhood cancer survivors: a report from the childhood cancer survivor study. Cancer Epidemiol Biomarkers Prev 2008;17:435-446.
9. Zebrack BJ, Zeltzer LK. Quality of life issues and cancer survivorship. Curr Probl Cancer 2003;27:198-211.
10. Breetvelt IS, Van Dam FS. Underreporting by cancer patients: the case of response-shift. Soc Sci Med 1991;32:981-987.
11. Apajasalo M, Sintonen H, Siimes MA, et al. Health-related quality of life of adults surviving malignancies in childhood. Eur J Cancer 1996;32A:1354-1358.
12. Sundberg KK, Lampic C, Bjork O, et al. Positive and negative consequences of childhood cancer influencing the lives of young adults. Eur J Oncol Nurs 2009;13:164-170.
13. Phipps S, Steele RG, Hall K, et al. Repressive adaptation in children with cancer: a replication and extension. Health Psychol 2001;20:445-451.
14. O'Leary TE, Diller L, Recklitis CJ. The effects of response bias on self-reported quality of life among childhood cancer survivors. Qual Life Res 2007;16:1211-1220.
15. Verhoof E, Maurice-Stam H, Heymans H, et al. Growing into disability benefits? Psychosocial course of life of young adults with a chronic somatic disease or disability. Acta Paediatr 2011.
16. Madan-Swain A, Brown RT, Sexson SB, et al. Adolescent cancer survivors. Psychosocial and familial adaptation. Psychosomatics 1994;35:453-459.

17. Oeffinger KC, Mertens AC, Sklar CA, et al. Chronic health conditions in adult survivors of childhood cancer. N Engl J Med 2006;355:1572-1582.
18. Blaauwbroek R, Stant AD, Groenier KH, et al. Health-related quality of life and adverse late effects in adult (very) long-term childhood cancer survivors. Eur J Cancer 2007;43:122-130.
19. Zebrack BJ, Stuber ML, Meeske KA, et al. Perceived positive impact of cancer among long-term survivors of childhood cancer: a report from the childhood cancer survivor study. Psychooncology 2011.
20. Teresi JA. Different approaches to differential item functioning in health applications. Advantages, disadvantages and some neglected topics. Med Care 2006;44:S152-70.
21. Teresi JA, Fleishman JA. Differential item functioning and health assessment. Qual Life Res 2007;16 (Suppl 1):33-42.
22. Reulen RC, Zeegers MP, Jenkinson C, et al. The use of the SF-36 questionnaire in adult survivors of childhood cancer: evaluation of data quality, score reliability, and scaling assumptions. Health Qual Life Outcomes 2006;4:77.
23. Huang IC, Quinn GP, Eddleton KZ, et al. Head-to-head comparisons of quality of life instruments for young adult survivors of childhood cancer. Support Care Cancer 2011 (In Press; DOI: 10.1007/s00520-011-1315-5).
24. Huang IC, Quinn GP, Wen PS, et al. Developing item banks of health-related quality of life measurement for young adult survivors of childhood cancer. Qual Life Res 2011 (In Press; DOI: 10.1007/s11136-011-0055-9).
25. Jones RN. Identification of measurement differences between English and Spanish language versions of the Mini-Mental State Examination. Detecting differential item functioning using MIMIC modeling. Med Care 2006;44:S124-33.
26. Carle AC. Mitigating systematic measurement error in comparative effectiveness research in heterogeneous populations. Med Care 2010;48:S68-74.
27. Recklitis CJ, Parsons SK, Shih MC, et al. Factor structure of the brief symptom inventory--18 in adult survivors of childhood cancer: results from the childhood cancer survivor study. Psychol Assess 2006;18:22-32.
28. Stuber ML, Meeske KA, Leisenring W, et al. Defining medical posttraumatic stress among young adult survivors in the Childhood Cancer Survivor Study. Gen Hosp Psychiatry 2011;33:347-353.
29. Yang FM, Tommet D, Jones RN. Disparities in self-reported geriatric depressive symptoms due to sociodemographic differences: an extension of the bi-factor item response theory model for use in differential item functioning. J Psychiatr Res 2009;43:1025-1035.
30. Cohen J. Statistical power analysis for the behavioral sciences. Hillsdale, N.J.: L. Erlbaum Associates; 1988.

APPENDIX: SAMPLE TABLES AND FIGURES

Table 1: Characteristics of study population

	YASCC (N=)	Sibling (N=)
Age, (mean, SD) [in year]		
Sex, (N, %)		
Male		
Female		
Race/ethnicity, (N, %)		
White, non-Hispanic		
Black, non-Hispanic		
Hispanic		
Other		
Educational background, (N, %)		
Below high school		
High school graduate/ GED		
Some college/ training after high school		
College graduate		
Post graduate level		
Marital status, (N, %)		
Married/ living with a partner		
Widowed/ divorced/ separated		
Single		
Employment status, (N, %)		
Working full-time		
Working part-time		
Others (will breakdown depending upon frequency)		
Insurance status, (N, %)		
Insured		
Uninsured		
Annual household incomes, (N, %)		
< $19,999		
$20,000 – $39,999		
$40,000 – $59,999		
$60,000 – $79,999		
$80,000 – $99,999		
≥ $100,000		
Age at diagnosis, (mean, SD) [in year]		
Age at interview, (mean, SD) [in year]		
Time since diagnosis, (mean, SD) [in year]		
Cancer diagnosis, (N, %)		
Leukemia		
Central nervous system (CNS) tumor		
Hodgkin lymphoma		

Non-Hodgkin lymphoma		
Wilms tumor		
Neuroplastoma		
Soft tissue sarcoma		
Bone tumor		
Second cancer, (N, %)		
Yes		
No		
Chemotherapy, (N, %)		
None		
Methotrexate		
Corticosteroid		
Anthracyclines		
Alkylating agents		
Other		
Radiotherapy, (N, %)		
None		
Cranial		
Other		
Surgery, (N, %)		
None		
Amputation		
Other		
Severity of chronic condition by CTCAE, (N, %)		
Grade 1		
Grade 2		
Grade 3		
Grade 4		
Body mass index (BMI) (N, %)		
Underweight (<18.5 kg/m^2)		
Normal weight (18.5 – 24.9 kg/m^2)		
Overweight (25.0 – 29.9 kg/m^2)		
Obese (≥30 kg/m^2)		
Brief Symptom Inventory (N, %)		
Anxiety (cutoff: 63)		
Depression (cutoff: 63)		
Somatization (cutoff: 63)		
Global severity index (cutoff: 63)		
Post-Traumatic Stress Disorder (N, %)		
Yes (DSM IV diagnostic criteria)		
No		
Perceived Positive Impact (N, %)		
Yes (cutoff derived from sibling)		
No		

Table 2: Parameters of SF-36 items for YASCC and Siblings

	YASCC vs. siblings	Mean raw item score	Difference in effect size*	Item parameters (standard error)† from the final model				
				Factor loading	1st threshold	2nd threshold	3rd threshold	4th threshold
Domain 1 of the SF-36								
Item 1	YASCC							
	Siblings							
Item 2	YASCC							
	Siblings							
Item 3	YASCC							
	Siblings							
Domain 2 of the SF-36								
Item 1	YASCC							
	Siblings							
Item 2	YASCC							
	Siblings							
Item 3	YASCC							
	Siblings							
Domain 3 of the SF-36								
Item 1	YASCC							
	Siblings							
Item 2	YASCC							
	Siblings							
Item 3	YASCC							
	Siblings							
Domain 4 of the SF-36								
Item 1	YASCC							
	Siblings							
Item 2	YASCC							
	Siblings							
Item 3	YASCC							
	Siblings							

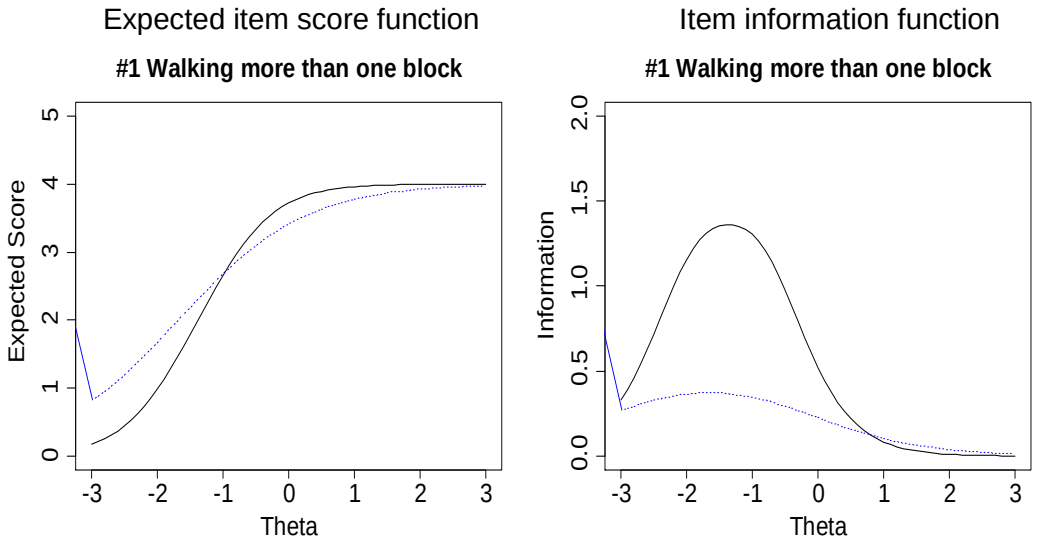
* Magnitude of effect size (in an absolute value): negligible: <0.2, small: 0.2-0.49, moderate: 0.5-0.79, and large: ≥ 0.8.
† If an item was identified with DIF, different values of a parameter were presented in both groups.
‡ $p < 0.001$.

Table 3: Underlying HRQOL scores between YASCC and siblings before and after DIF calibration

	Before DIF calibration	After DIF calibration	Subjects who increase or decrease scores by ≥ 2 SD, respectively, after calibration (%)
Domain 1 of the SF-36			
YASCC			
Siblings			
Difference (effect size)*,†			
Domain 2 of the SF-36			
YASCC			
Siblings			
Difference (effect size) *,†			
Domain 3 of the SF-36			
YASCC			
Siblings			
Difference (effect size) *,†			
Domain 4 of the SF-36			
YASCC			
Siblings			
Difference (effect size) *,†			

* Magnitude of effect size (in an absolute value): negligible: <0.2, small: 0.2-0.49, moderate: 0.5-0.79, and large: ≥ 0.8.
† $p < 0.001$.

Figure 2: Expected item score function and item information function of DIF items (hypothetical examples)



Note, solid line: YASCC; dotted line: Siblings