Floor and Ceiling Effects in the EORTC QLQ-C30 Physical Functioning Subscale Among Patients With Advanced or Metastatic Breast Cancer

Meena N. Murugappan, PharmD, MPH, PhD(c) [1]; Bellinda L. King-Kallimanis, PhD[2]; Christina Mangir, MS[3]; Lynn Howie, MD[4]; Vishal Bhatnagar, MD[2]; Julia A. Beaver, MD[2]; Ethan M. Basch, MD, MSc [5]; Sydney R. Henson, BS[5]; and Paul G. Kluetz, MD[2]

BACKGROUND: The European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 Physical Functioning subscale is a widely used patient-r eported outcome measure that quantifies cancer patients' physical functioning. Strong floor/ceiling effects can affect a scale's sensitivity to change. The aim of this study was to characterize floor/ceiling effects of the physical functioning domain in patients with advanced/metastatic breast cancer enrolled in commercial clinical trials and a community-based trial. **METHODS:** The clinical trial cohort comprised patients from 5 registrational trials submitted to the Food and Drug Administration for review (2010-2017). The community cohort comprised a subgroup of patients from the Alliance Patient Reported Outcomes to Enhance Cancer Treatment (PRO-TECT) trial. The distribution of patient responses to Physical Functioning items and the summed score were assessed at the baseline and 3-month follow-up for both cohorts. Descriptive statistics were used to determine floor/ceiling effects at the item and scale levels. **RESULTS:** The clinical trial cohort and the community cohort consisted of 2407 and 178 patients, respectively. Twenty-f our percent or more of the respondents reported "not at all" for having trouble/needing help with each Physical Functioning item across both cohorts and measurement time points. Fourteen to twenty percent of the patients scored perfectly (100 of 100) on the Physical Functioning subscale summary measure (where higher scores indicated better physical functioning) across both cohorts and time points. **CONCLUSIONS:** Minor floor effects and notable ceiling effects were found at the item and scale levels of the Physical Functioning subscale, regardless of cohort, and this creates some uncertainty about its ability to detect changes in physical functioning among high-functioning patients. Investigators may consider adding additional high-functioning items from the EORTC's item library to more accurately describe the impact of anticancer treatment on patients' physical functioning. ***Cancer* 2022;128:808-818**.

KEYWORDS: breast neoplasms, European Organization for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 (EORTC QLQ-C30), patient-reported outcome measures, physical functioning, quality of life.

INTRODUCTION

Patient-reported outcome (PRO) measures are a standardized method of collecting information on a patient's health status that comes directly from the patient without interpretation by a clinician or anyone else.[1] These measures can provide valuable information on a patient's disease symptoms, treatment side effects, and physical functioning as well as other aspects of his or her health-related quality of life (HRQL). When rigorously assessed with clear objectives, PRO end points can be useful for regulatory and clinical decision-making.[2,3] Together, clinical and PRO data can provide a comprehensive assessment of the impact of a disease and its treatment on the daily life of patients with cancer.

Key concerns raised by patients with breast cancer during a Food and Drug Administration (FDA) Patient-Focused Drug Development Listening Session were the limitations that they experienced in physical functioning and activity as a result of their disease and treatment.[4] Many of the trials evaluating novel agents for advanced/metastatic breast cancer submitted to the FDA include PRO measures that capture physical functioning. One of the most commonly used measures is the European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-C30).[5] This general cancer questionnaire consists of 30 items, 24 of which are aggregated into 9 multi-item scales: 1 global health status scale, 3 symptom scales, and 5 functional scales (physical, role, cognitive, social, and emotional). The Physical Functioning subscale consists of 5 items asking respondents about the level of help that they need

Corresponding Author: Meena N. Murugappan, PharmD, MPH, PhD(c), Oncology Center of Excellence, US Food and Drug Administration, 10903 New Hampshire Ave, Bldg 22, Silver Spring, MD 20993 (meena.murugappan@fda.hhs.gov).

[1]Office of Oncologic Diseases, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, Maryland; [2]Oncology Center of Excellence, US Food and Drug Administration, Silver Spring, Maryland; [3]Association of Community Cancer Centers, Rockville, Maryland; [4]Pardee Hospital, University of North Carolina Health, Hendersonville, North Carolina; [5]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina

The first 2 authors contributed equally to this article.

DOI: 10.1002/cncr.33959, **Received:** May 14, 2021; **Revised:** July 19, 2021; **Accepted:** August 16, 2021, **Published online** October 11, 2021 in Wiley Online Library (wileyonlinelibrary.com)

or the level of trouble that they have with various physical activities such as carrying a heavy shopping bag or taking a short walk.[6,7] Although the reliability and validity of the QLQ-C30 have been exhaustively investigated across cancer types and in culturally diverse patient populations,[8-12] less is known regarding the Physical Functioning subscale specifically and its responsiveness in the current advanced/metastatic breast cancer treatment landscape.

Responsiveness refers to the degree to which a scale can demonstrate clinical deterioration and improvement. It can be affected by the proportion of respondents who report the worst/minimum or best/maximum score on the scale; these are also known as floor and ceiling effects, respectively.[6] A floor effect is defined as the clustering of participants' responses toward the bottom end or worst possible score of a scale or instrument.[13] A ceiling effect is the opposite.[13] For an item with 4 to 5 response options, a strong floor effect occurs when a considerable proportion of respondents (eg, >20%) respond with the worst response option (eg, reporting "very much" difficulty with a given physical activity).[14] Meanwhile, a ceiling effect occurs when >20% of the respondents select the best response option (eg, reporting "not at all" for difficulty with a given physical activity).[14] Floor and ceiling effects can influence a scale's responsiveness to change because they limit our ability to measure variance above or below a certain limit.[15] Although some degree of floor and ceiling effects is expected for any measure, these effects make it difficult to distinguish among study participants at the top or bottom end of a scale.[16] They are an indication that the Physical Functioning items are not challenging enough in the case of strong ceiling effects and too challenging in the case of strong floor effects. When a large proportion of items in a scale display the ceiling effect, it may be interpreted as evidence of the instrument's inability to discriminate among the highest levels of a construct such as physical functioning.[16] Strong ceiling effects thus hinder our ability to detect improvement after an intervention or reveal deterioration over time. To reliably measure the full spectrum of physical functioning in advanced/metastatic breast cancer, the items used should match the functional ability of the study population.[17] One indicator that a scale may not be responsive for a given patient population is the presence of pronounced floor and/or ceiling effects.

In their systematic review, Luckett et al[6] found that substantial floor and ceiling effects were common among studies investigating the responsiveness of the QLQ-C30 in patients with different types of cancers. However, there is a paucity of literature regarding the responsiveness of the QLQ-C30 in patients with breast cancer specifically, and this handful of studies has reported conflicting results. Demirci et al[9] investigated floor/ceiling effects at the scale level in a cohort of patients with breast cancer in the adjuvant treatment setting. They found that 0% of the study participants scored 0 of 100 and 5.5% scored 100 of 100 on the Physical Functioning subscale. Meanwhile, Alawadhi and Ohaeri[11] assessed floor/ceiling effects of the Physical Functioning subscale at the item level in women with advanced breast cancer (stages III and IV). Floor effects for items in their study ranged from 3.3% to 14.4%, whereas ceiling effects ranged from 7.5% to 20.9% (depending on the item).[11] Importantly, these studies were intended to validate the EORTC QLQ-C30 in different cultural and geographic contexts. Their primary aim was not to assess the responsiveness of the measure over the course of patients' treatment. These studies were also limited by their cross-sectional nature, small sample sizes, and ambiguity regarding the time that elapsed between the receipt of treatment and the administration of the QLQ-C30 questionnaire.[9,11] Therefore, an opportunity exists to evaluate floor and ceiling effects of the EORTC QLQ-C30 Physical Functioning subscale in a larger cohort of patients receiving treatment for advanced/metastatic breast cancer.

The aims of this study were to 1) assess the EORTC QLQ-C30 Physical Functioning subscale and individual items for floor and ceiling effects when administered to patients receiving pharmacological treatment for advanced/metastatic breast cancer and 2) compare Physical Functioning subscale floor/ceiling effects between patients enrolled in commercial clinical trials and patients who were being treated in the community as part of standard clinical care. Findings from this study could help to inform the interpretation of clinical trial results in trials using the QLQ-C30. Study findings could also identify opportunities for developing responsive and relevant methods for determining changes in patients' physical functioning over the course of cancer treatment.

MATERIALS AND METHODS

Data Source: The FDA Clinical Trial Cohort

US FDA internal databases were searched to identify breast cancer registration trials that incorporated the EORTC QLQ-C30 questionnaire. Trials were included in this analysis if they were submitted to the FDA between 2010 and 2017 and supported FDA approval for the treatment of advanced/metastatic breast cancer. Five trials were identified as meeting the inclusion/exclusion criteria: 1 single-arm trial and 4 double-blind randomized trials (MONARCH-1, MONARCH-2, MONARCH-3, PALOMA-3, and

MONALEESA-2). All 5 trials involved oral targeted treatments (CDK4/6 inhibitors) as single agents or in combination with hormone therapy. To be included in these trials, patients had to have an Eastern Cooperative Oncology Group (ECOG) performance status score of 0 or 1 and no laboratory findings indicative of organ dysfunction. Baseline PRO data (from within 30 days before treatment initiation) were pooled and evaluated. Patients were included in this analysis only if they completed all 5 items of the Physical Functioning subscale at the baseline because this was essential to assess any potential changes in physical functioning over time. For the follow-up PRO assessment, we focused on day 1 of cycle 3 because all trials had administered the EORTC QLQ-C30 at this visit. The FDA Center for Drug Evaluation and Research Human Subject Protection Liaison to the FDA Institutional Review Board (IRB) determined that this study is consistent with a "not human subject research" determination and thus does not require IRB review.

Data Source: The Community Cohort

The Alliance Patient Reported Outcomes to Enhance Cancer Treatment (PRO-TECT) trial (NCT03249090) was designed to assess symptom management in a community-based cohort of patients. PRO-TECT was not a therapeutic intervention trial and, therefore, better reflected standard clinical care. This trial also included the EORTC QLQ-C30 questionnaire. Our study involved a secondary evaluation of a subgroup of patients from the original trial data set: those who were actively receiving treatment for advanced/metastatic breast cancer and had completed the QLQ-C30 at the baseline (ie, before randomization). For follow-up, we used patient responses from their month 3 PRO assessment.

Instruments

We focused on the 5 items from the QLQ-C30 that measure physical functioning (ie, the Physical Functioning subscale). These items range from "trouble doing strenuous activities" to "help with eating, dressing, washing yourself, and using the toilet". Items are measured with a 4-point response option scale ("not at all," "a little," "quite a bit," and "very much"). Responses are then reversed, summed, and transformed to range from 0 to 100 to create a summary score. Higher summary scores indicate better physical functioning. Scores in this study were calculated according to the official scoring algorithms in the EORTC QLQ-C30 scoring manual.

Study Design and Statistical Analysis

Descriptive statistics were used to summarize patients' demographic and clinical characteristics at the baseline. Floor/ceiling effects and the distribution of patient responses for each of the 5 individual items of the Physical Functioning subscale were evaluated at the baseline and follow-up for both cohorts (clinical trial and community). For the purposes of this study, a strong floor effect at the item level was defined as >20% of the patients reporting the worst possible score or level of physical functioning (ie, reporting "very much" difficulty with a given item/physical activity). Meanwhile, a strong ceiling effect at the item level was defined as >20% of the patients reporting the best possible score or level of physical functioning (ie, reporting "not at all" for difficulty with a given item/physical activity).

Next, floor and ceiling effects for the summary measure of the Physical Functioning subscale were calculated for both cohorts at both time points. For the purposes of this study, a strong floor effect at the scale level was defined as >20% of the patients scoring ≤6.7 out of 100 on the summary measure. A cutoff of ≤6.7 was selected because this reflected patients who reported a minimum of "quite a bit" of difficulty on all 5 of the Physical Functioning items. Namely, a patient score of 0 on the summary measure would mean that they reported "very much" difficulty with all 5 items. Alternatively, a score of 6.7 would mean that they reported "quite a bit" of difficulty on 1 of the items and "very much" difficulty on the remaining 4 items. Because the majority of the items, if not all, were answered with the worst response option, these patients would meet our study's definition of experiencing a strong floor effect at the scale level (score of ≤6.7 on the summary measure).

Similarly, for the purposes of this study, a ceiling effect at the scale level was defined as >20% of the patients scoring ≥93.3 out of 100 on the summary scale. A cutoff of ≥93.3 was selected because this reflected that these patients reported a maximum of "a little" difficulty on all 5 of the Physical Functioning items. That is, a patient score of 100 on the summary measure would mean that the patient reported "not at all" for difficulty with all of the 5 items. Alternatively, a score of 93.3 would mean that the patient reported "a little" of difficulty on just 1 of the items and "not at all" for difficulty on the remaining 4 items. Because the majority of the items, if not all, were answered with the best response option, these patients would meet our study's definition of experiencing a strong ceiling effect at the scale level (score of ≥93.3 on the summary measure).

Changes in patients' responses to the Physical Functioning domain were depicted with Sankey diagrams. Patients who did not complete their PRO assessment at follow-up were included in this analysis but were grouped together as having "no assessment" at follow-up.

TABLE 1. Distribution of Sociodemographic and Clinical Characteristics in the Clinical Trial Cohort (n = 2407) and the Community Cohort (n = 178)

Variable	Clinical Trial Cohort, No. (%)	Community Cohort, No. (%)
Sociodemographics		
Age group		
≤40 y	101 (4.2)	12 (6.7)
41-50 y	371 (15.4)	34 (19.1)
51-60 y	727 (30.2)	63 (35.4)
>60 y	1208 (50.2)	69 (38.8)
Gender		
Female	2407 (100)	173 (97.2)
Race		
White	1120 (46.5)	136 (76.4)
Asian	531 (22.1)	2 (1.1)
Native American	33 (1.4)	5 (2.8)
Black	59 (2.5)	30 (16.9)
Other	148 (6.1)	5 (2.8)
Ethnicity		
Hispanic or Latino	N/A	8 (4.5)
Non-Hispanic	N/A	170 (95.5)
Geographic region		
United States	550 (22.9)	N/A
Outside United States	1857 (77.1)	N/A
Education level		
1st to 11th grades	N/A	9 (5.1)
High school graduate/ GED	N/A	55 (30.9)
Some college/associate's degree	N/A	67 (37.6)
Bachelor's degree	N/A	28 (15.7)
Advanced degree (MA, PhD, etc)	N/A	16 (9)
Other	N/A	3 (1.7)
Employment status		
Full time (≥40 h/wk)	N/A	38 (21.3)
Part time	N/A	29 (16.3)
Not currently working	N/A	111 (62.4)
Marital status		
Single, never married	N/A	20 (11.2)
Married/partnered	N/A	109 (61.2)
Separated/divorced	N/A	28 (15.7)
Widowed	N/A	21 (11.8)
Clinical characteristics		
Current line of treatment		
1st line	1095 (45.5)	45 (25.3)
2nd line or later	1312 (54.5)	133 (74.7)
Baseline ECOG score		
0: fully active	1455 (60.4)	78 (44.3)
1	948 (39.5)	79 (44.9)
2	0	18 (10.2)
3	0 (0)	1 (0.6)
4: completely disabled	0 (0)	0 (0)
Missing	4 (<1.0)	N/A
Year of diagnosis		
1985-1995	N/A	4 (2.2)
1996-2006	N/A	27 (15.2)
2007-2018	N/A	139 (78.1)
Year of metastasis		
2003-2008	N/A	6 (3.4)
2009-2013	N/A	21 (11.8)
2014-2018	N/A	137 (77)
Dosage form of treatment		
Receiving IV treatment only	0	129 (72.5)
Receiving PO treatment only	1900 (78.9)	35 (19.7)
Receiving PO + IV treatment	N/A	14 (7.9)
Receiving PO + IM treatment	512 (20.9)	N/A
Randomized, not treated	5 (<1.0)	N/A

Abbreviations: ECOG, Eastern Cooperative Oncology Group; GED, General Educational Development; IM, intramuscular; IV, intravenous; N/A, not available; PO, oral administration.

All statistical analyses were performed with SAS (version 9.4) and RStudio (version 1.2.5001). Sankey diagrams were generated with the networkD3 package in RStudio.

RESULTS

Baseline Characteristics of the Clinical Trial and Community Cohort Patients

The clinical trial cohort and community cohort consisted of 2407 and 178 patients, respectively (Table 1). Most patients in both cohorts were female, older than 50 years, and White. They were from varied metastatic treatment settings (first line and beyond), and all were receiving oral targeted treatments with or without hormone therapy, which was administered either orally or by intramuscular injection. Although 60% of the clinical trial patients had a baseline ECOG performance score of 0 (fully active), only 44% of the patients in the community cohort had a baseline ECOG score of 0. Additional demographic and clinical information was available only for community-based patients; approximately two-thirds were married and not employed at the time of the study. Most were receiving intravenous therapy, and at least half of this cohort were on their third or fourth line of treatment.

Item-Level Floor Effects in the Clinical Trial Cohort Patients Versus the Community Cohort Patients

Floor effects for Physical Functioning items ranged from 1% to 13% at the baseline and from 1% to 11% at follow-up in the clinical trial cohort. Floor effects were higher in the community cohort and ranged from 1% to 23% at the baseline and from 2% to 18% at follow-up (Table 2).

Item-Level Ceiling Effects in the Clinical Trial Cohort Patients Versus the Community Cohort Patients

Item-level ceiling effects at the baseline and follow-up (ie, the percentages of patients who responded "not at all" to whether they had difficulty with a task) are presented in

TABLE 2. Baseline and Follow-Up Item-Level Floor and Ceiling Effects in Clinical Trial Cohort and Community Cohort Patients

Item	Question	Clinical Trial Cohort, % (No.)		Community Cohort, % (No.)	
		Baseline (n = 2407)	Cycle 3 (n = 1982)	Baseline (n = 178)	Month 3 (n = 159)
Item-Level Floor Effects: % of Patients Responding "Very Much" (Low Physical Functioning)					
1	Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?	13 (324)	11 (221)	20 (36)	15 (24)
2	Do you have any trouble taking a long walk?	12 (300)	11 (219)	23 (41)	18 (29)
3	Do you have any trouble taking a short walk outside of the house?	2 (59)	2 (43)	8 (15)	4 (7)
4	Do you need to stay in bed or a chair during the day?	3 (66)	2 (47)	5 (9)	9 (14)
5	Do you need help with eating, dressing, washing yourself, or using the toilet?	1 (18)	1 (10)	1 (1)	2 (3)
Item-Level Ceiling Effects: % of Patients Responding "Not At All" (High Physical Functioning)					
1	Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?	27 (646)	28 (563)	24 (42)	27 (43)
2	Do you have any trouble taking a long walk?	34 (827)	29 (577)	24 (42)	30 (47)
3	Do you have any trouble taking a short walk outside of the house?	71 (1715)	69 (1374)	58 (104)	64 (101)
4	Do you need to stay in bed or a chair during the day?	60 (1446)	58 (1159)	46 (82)	51 (81)
5	Do you need help with eating, dressing, washing yourself, or using the toilet?	94 (2273)	94 (1871)	89 (159)	92 (146)

The difference in denominators between the baseline and cycle 3/month 3 is due to fewer patients completing the Quality of Life Questionnaire Core 30 during their follow-up visit.

Table 2. Large ceiling effects were observed at the baseline and follow-up in both clinical trial and community-based patients. Ceiling effects were somewhat more pronounced in the clinical trial cohort versus the community cohort. For example, at the baseline, 34% of the clinical trial patients reported that they had no trouble at all when taking a long walk (item 2 of the Physical Functioning subscale), whereas 24% of the patients in the community cohort reported that they had no trouble at all for the same item. The proportion of clinical trial patients responding "not at all" to each question on the Physical Functioning subscale was slightly less during the follow-up visit in comparison with the baseline (Fig. 1). On the other hand, ceiling effects in the community cohort were slightly elevated at follow-up in comparison with the baseline, although this may be due to a decreased number of respondents at follow-up (159 at follow-up vs 178 at the baseline). Ceiling effects were especially noticeable, regardless of the measurement time point or cohort, for the final 3 items of the Physical Functioning subscale: trouble with a short walk, staying in bed or a chair all day, and needing help with eating.

Scale-Level Floor Effects in the Clinical Trial Cohort Patients Versus the Community Cohort Patients

Floor effects at the scale level, reflecting participants who scored ≤6.7 on the summary measure, were negligible (<1%) both at the baseline and at follow-up in the clinical trial cohort. On the contrary, although <1% of the patients in the community cohort scored ≤6.7 on the summary measure at the baseline, 14% scored ≤6.7 at follow-up (Table 3).

Scale-Level Ceiling Effects in the Clinical Trial Cohort Patients Versus the Community Cohort Patients

Thirty percent of clinical trial patients and 26% of community-based patients scored ≥93.3 on the Physical Functioning subscale at the baseline (Table 3). This proportion increased slightly at follow-up: 32% of clinical trial patients and 30% of community-based patients scored ≥93.3 on the summary measure. Little change in ceiling effects was observed over time, regardless of the cohort (Fig. 2).

Figure 1. Distribution of patient responses at the baseline and at follow-up for each item of the Physical Functioning subscale. These Sankey diagrams depict how patients' responses to each item of the Physical Functioning subscale changed over time. For example, in item 1, in the clinical trial cohort, 27% of all patients at the baseline reported that they had no trouble at all with doing strenuous activities, 38% reported having a little trouble, 23% had quite a bit of trouble, and 13% had very much trouble. Eighteen of the patients who took the Quality of Life Questionnaire Core 30 at the baseline did not take it at cycle 3 (no assessment). The overall percentage of patients responding "not at all," "a little," "quite a bit," and "very much" remained relatively stable between the baseline and follow-up, regardless of the cohort. For example, the percentage of patients responding "not at all" to item 1 went from 26% at the baseline to 23% at the cycle 3 follow-up in the clinical trial cohort. Only a small percentage of patients (as seen in the colored links) transitioned from one response level at the baseline to another at follow-up. Similar patterns can be observed in the Sankey diagrams for patients in the community cohort.

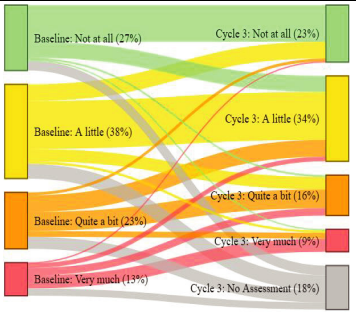
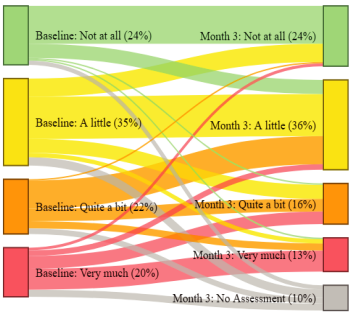
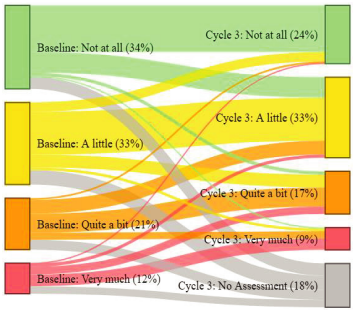
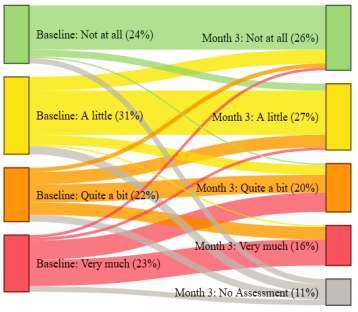
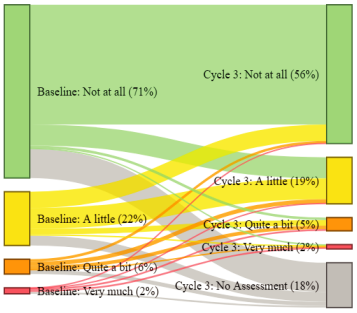
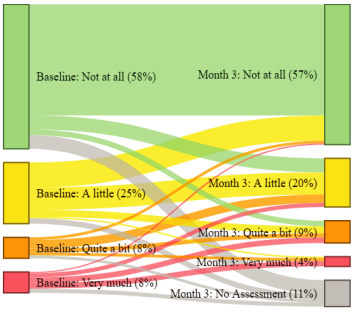
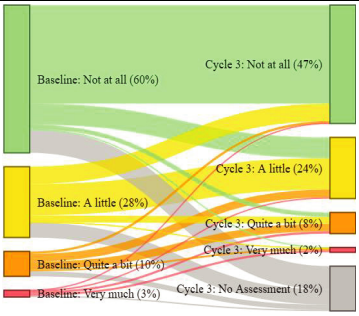
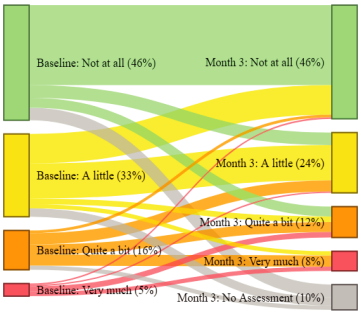
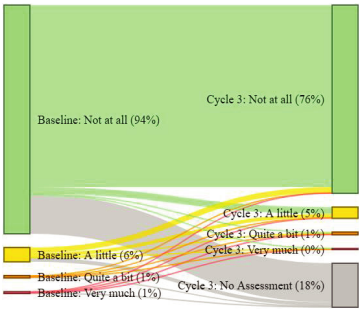
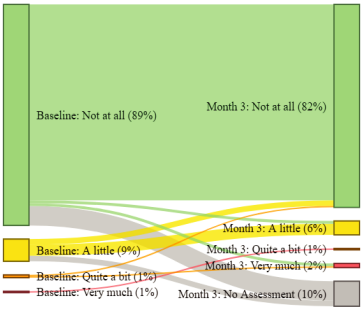
Clinical Trial Cohort (N = 2407)	Community Cohort (N = 178)
A Item 1: Do you have any trouble doing strenuous activities, like carrying a heavy shopping bag or a suitcase?	
	
B Item 2: Do you have any trouble taking a long walk?	
	
C Item 3: Do you have any trouble taking a short walk outside of the house?	
	
D Item 4: Do you need to stay in bed or a chair during the day?	
	
E Item 5: Do you need help with eating, dressing, washing yourself or using the toilet?	
	

TABLE 3. Baseline and Follow-Up Scale-Level Floor and Ceiling Effects in Clinical Trial Cohort Patients and Community Cohort Patients

	Clinical Trial Cohort, % (No.)		Community Cohort, % (No.)	
	Baseline (n = 2407)	Cycle 3 (n = 1982)	Baseline (n = 178)	Month 3 (n = 159)
Scale-Level Floor Effects: % of Patients Scoring ≤6.7/100 on the 100-Point Physical Functioning Subscale Summary Measure (Low Physical Functioning)				
Patients scoring 0 (out of 100)	0.2 (5)	0.3 (6)	0.5 (1)	13 (20)
Patients scoring 6.7 (out of 100)	0.3 (8)	0.2 (3)	0 (0)	0.6 (1)
Total	0.5 (13)	0.5 (9)	0.5 (1)	14 (21)
Scale-Level Ceiling Effects: % of Patients Scoring ≥93.3/100 on the 100-Point Physical Functioning Subscale Summary Measure (High Physical Functioning)				
Patients scoring 93.3 (out of 100)	14 (348)	13 (261)	13 (23)	11 (17)
Patients scoring 100 (out of 100)	18 (438)	17 (346)	14 (24)	20 (31)
Total	32 (786)	30 (607)	26 (47)	30 (48)

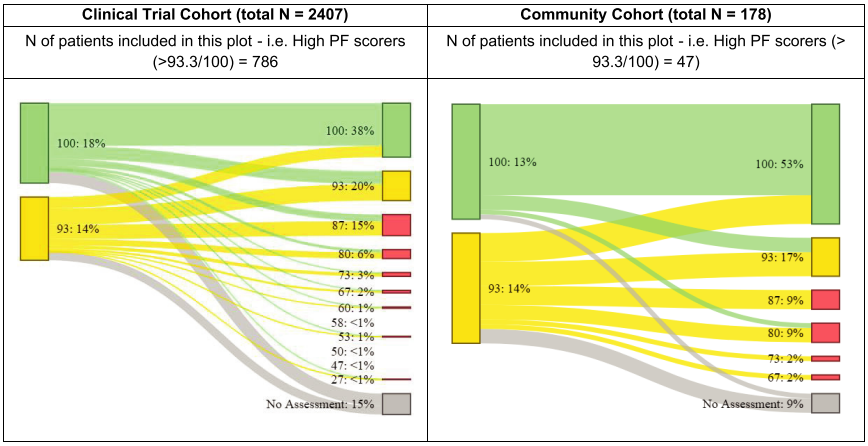


Figure 2. Sankey plots of the ceiling effects observed in the PF subscale summary score. These Sankey diagrams illustrate how high-functioning patients' PF summary scores changed over time. Higher scores indicate better physical functioning. We limited these diagrams to solely include patients who were high functioning at the baseline (ie, they scored 93.3 or higher on the PF subscale). This means that these patients reported "not at all" on (almost) all PF items. As shown in the diagram on the left, 18% of the clinical trial cohort patients scored 100 out of 100 on the PF subscale at the baseline, whereas 38% of all clinical trial cohort patients who took the Quality of Life Questionnaire Core 30 at cycle 3 scored 100 out of 100. Thirteen percent of community-based patients scored 100 at the baseline, whereas 53% scored 100 at their month 3 follow-up visit, as shown in the diagram on the right. However, a greater proportion of patients scoring perfectly at follow-up is not reflective of an improvement in physical functioning over the course of treatment. Rather, it is a function of fewer patients taking the patient-reported outcome assessment at follow-up as well as some improvement in physical functioning, as represented by the yellow links flowing from a score of 93 at the baseline to a score of 100 at follow-up. It is equally important to note that there were some patients whose physical functioning declined from the baseline to follow-up; they are represented by the green links that flow from 100 at the baseline to lower scores at follow-up. A key takeaway from this pair of Sankey diagrams is that the PF subscale captured little to no change among the highest functioning patients across both cohorts; the majority of the patients who started with a score of 93.3 or higher remained there at follow-up. PF indicates Physical Functioning.

DISCUSSION

In this study of more than 2400 patients with advanced/metastatic breast cancer, we found that the Physical Functioning domain of the EORTC QLQ-C30 was associated with minor floor effects and notable ceiling effects at the item and scale levels, regardless of whether patients were enrolled in commercial clinical trials or were being treated at their community cancer center. Significant ceiling effects at the item level indicate that a substantial proportion of patients felt that they had no trouble doing most of the activities assessed by the scale at the baseline and follow-up. Likewise, our finding of a ceiling effect at the scale level, where higher scores indicate better physical functioning, demonstrates that a large proportion of respondents were clustered at the higher end of the Physical Functioning spectrum. This suggests that the Physical Functioning domain may not be sensitive enough to detect changes in physical functioning among patients who are high functioning at the baseline. For example, patients who were able to run/jog 3 miles/5 kilometers before treatment initiation but can no longer do so during or after treatment may still report that they

have no trouble with activities such as carrying a heavy shopping bag (item 1) or taking a long walk (item 2). The QLQ-C30 would not be able to capture this change in patients' physical functioning because it does not include items targeted at patients who may have been highly active and able at the baseline. Therefore, when the Physical Functioning subscale of the QLQ-C30 is used, there is chance that a false conclusion could be made for patients who are very high functioning (ie, there was no change in their physical functioning when in reality the patients may have experienced a decline).

A key finding from our study is that ceiling effects were most noticeable in the final 3 items of the Physical Functioning subscale, as evidenced by our baseline results, where 51% to 94% of the patients reported no trouble at all with activities such as eating and going for a short walk. This finding is not surprising because the QLQ-C30's Physical Functioning subscale is a Guttman scale in which respondents who report little difficulty with the first item are unlikely to report serious difficulty with any of the subsequent items.[6] Additionally, we found that ceiling effects were more pronounced among patients who were enrolled in clinical trials rather than their community-based counterparts. This was expected because clinical trial patients are selected via a set of stringent inclusion/exclusion criteria. Indeed, a greater proportion of clinical trial patients in our study had a baseline ECOG performance score of 0 in comparison with their community-based counterparts. Another potential explanation for the observed difference in ceiling effects between the 2 cohorts is the fact that the community-based patients were more likely to be on intravenous therapy and on their third or fourth line of treatment. The increased toxicity associated with intravenous chemotherapy regimens and the more advanced stage of disease in the community cohort may have resulted in their lower levels of physical functioning and consequently less pronounced ceiling effects.

When significant ceiling effects are present, it can limit the responsiveness of the scale and blunt its sensitivity to change. This may be one reason for the results seen in our Sankey diagrams (Fig. 1), in which a considerable proportion of patients' responses do not change between the baseline and the cycle 3/month 3 follow-up. However, this should be interpreted with caution because physical functioning can be affected both positively (tumor-related symptom palliation) and negatively (toxicity) by cancer treatments; this makes it possible that in reality there was no change in physical functioning for a reasonable subset of patients, particularly in light of the short follow-up in this study.

One opportunity to improve sensitivity is to use customizable measures for physical functioning that are tailored to the patient population being studied. The EORTC QLQ-C30 has an item library available to those wishing to add additional items to the subscales.[18] Also, the Patient-Reported Outcomes Measurement Information System was developed to be used as an item bank, although a static short form can be created to measure a broader continuum of physical functioning.[19] Ideally, the inclusion of the appropriate patient population in the item selection process can increase the sensitivity of the QLQ-C30 Physical Functioning domain to the population being studied. Finally, it must be acknowledged that no static questionnaire can capture the full spectrum of physical functioning. Thus, computer adaptive testing (CAT)—for example, the CAT version of the EORTC QLQ-C30 or Patient-Reported Outcomes Measurement Information System—can be used to progressively select questions on physical functioning on the basis of each individual patient's prior responses.[20-22] A growing body of evidence suggests that this method is superior in terms of lower floor/ceiling effects, a greater ability to distinguish clinically significant changes, and decreased respondent burden.[23,24] Although CAT has benefits, this approach is logistically complicated to implement in commercial cancer clinical trials where the benefit in physical functioning is not a key secondary end point.

Another opportunity to increase the sensitivity of existing patient-reported physical functioning measures is to complement this information with sensor data. Wearable devices are becoming widely available and may improve the ability to detect changes in individuals with high daily activity levels. Activity levels alone may still be insufficient because there is a need to consider capturing more nuanced information regarding the level of effort and difficulty associated with various physical activities.[25] For instance, patients may be engaging in the same physical activities as they did before treatment out of necessity (eg, grocery shopping or caring for family), but it may be more difficult and take more effort than usual. In this scenario, although their activity levels have not changed, the difficulty and effort involved have changed, and this is meaningful information to patients. Although we acknowledge that comprehensive measurement of physical functioning will add complexity to clinical trials, patient-reported physical functioning data can be strengthened with complementary data from physical activity trackers,

particularly when comparative function is a key trial objective.

Although the EORTC QLQ-C30 Physical Functioning subscale may have limitations in detecting change at the highest levels of functioning, it should be noted that it is still an effective measure for mobility related to daily living activities such as eating, using the toilet, and going for a short walk, which are necessary to maintain independence. Preserving this basic level of functioning, especially in advanced/metastatic treatment settings where basic levels of functioning may be affected, is an important treatment goal, and the QLQ-C30 Physical Functioning subscale can be used to assess this goal. The absence of major floor effects in our study, especially in the final 3 items of the Physical Functioning subscale, suggests that the QLQ-C30 may be effective at measuring and discriminating among lower levels of physical functioning. The QLQ-C30 also has significant practical advantages for use in commercial clinical trials. It is the most commonly used instrument in trials submitted to the FDA for review and has been culturally and linguistically validated across a wide range of languages. This is necessary for global trials; less than a quarter of the clinical trial patients included in our study were located in the United States. Thus, our findings do not preclude use of the EORTC Physical Functioning domain to inform the risks and benefits of cancer therapies more broadly and may be appropriate for superiority objectives. However, our work does add uncertainty to claims of equivalence or noninferiority. As has been previously discussed, in instances where a cancer trial has an important study objective surrounding physical functioning, some of the limitations from floor and ceiling effects can be mitigated by adding high-functioning items from the item bank or by adding a second data source measuring activity through wearable devices.

Beyond the end points related to physical functioning on which commercial clinical trials focus, there is a need to differentiate between physical functioning associated with normal daily living activities (eg, walking, bathing, and carrying groceries) and more intense exercise-r elated physical activity. The latter has been shown to have a positive impact on HRQL and symptom management, such as alleviating pain, fatigue, and insomnia.[26] Studies have reported that engaging in moderate to vigorous aerobic exercise of at least 75 to 150 minutes per week plus resistance exercise twice a week may reduce the risk of mortality in patients with breast cancer by 40% to 50%.[27,28] The impact of exercise-related physical activity on HRQL and survival outcomes further underscores the need for sensitive and patient-centric measurements of physical functioning in patients during and after cancer treatment.

Our findings should be interpreted cautiously because there are several important limitations. There were only 178 community-based patients in this study, and the community sample predominantly received treatment via intravenous administration, whereas the patients in the clinical trial sample predominantly received their treatment via a daily oral pill. This limits the generalizability of our results. Although our community cohort of patients is closer to representing real-world patients than those enrolled in commercial clinical trials, they do not adequately represent the average breast cancer patient population. This is due to the inherent self-selection bias associated with participation in the PRO-TECT trial. It is possible that patients who chose to participate in this community care trial were also more physically active, as illustrated by the fact that 89% of the patients had an ECOG performance score of 0 or 1. Likewise, this study included only patients with advanced/metastatic breast cancer appropriate for a CDK4/6 regimen with or without hormone therapy, and they are not representative of all breast cancer patients, such as those in the adjuvant treatment setting, those with localized breast cancer, or those receiving intravenous chemotherapy in the metastatic setting. These different populations are likely to have varying degrees of baseline physical functioning and consequently varied floor/ceiling effects for the Physical Functioning domain. Future studies have an opportunity to investigate the responsiveness of the EORTC QLQ-C30 Physical Functioning subscale across different treatment settings.

Another limitation of this study is the short follow-up and varied schedule of assessments between the clinical trial and community cohorts. Postbaseline assessments for the clinical trial patients occurred at day 1 of cycle 3, that is, approximately 2 months after the baseline. Meanwhile, assessments for the community-based patients occurred 3 months after the baseline. This difference in the timing of assessments may have had an impact on patient responses and ultimately on floor/ceiling effects.

In conclusion, in our study of patients with advanced/metastatic breast cancer, the EORTC QLQ-C30's Physical Functioning domain demonstrated minor floor effects and notable ceiling effects in both clinical trial data and a community-based study. Ceiling effects persisted across different treatment lines and cohorts of patients, and limited change was observed from the baseline to follow-up with the Physical Functioning subscale. Our findings create some uncertainty in the ability of the QLQ-C30 to detect changes in physical functioning

among high-functioning patients. Nonetheless, the EORTC QLQ-C30's Physical Functioning subscale is widely used, has been translated across many languages, and remains an important tool for international clinical trials to describe functional impacts across advanced cancer contexts. Limitations in sensitivity can be mitigated by avoiding its use when the trial objective is to make a claim of noninferiority or equivalence or alternatively using the item library to add a higher functioning question to increase its sensitivity. In instances where the comparative benefit in physical function is a key trial objective, analyses should focus on superiority, and exploration of wearable devices may be another option for complementing PRO data. Further research, including the exploration of item banks, CAT, the appropriate timing of assessments, and the use of wearable device data, is warranted to optimize the assessment of physical functioning in cancer trials. Being able to accurately describe changes in physical functioning that are associated with disease and treatment can add value to our understanding of the risks and benefits of cancer treatments, and it is critical that we continue to advance rigorous methods to assess this outcome.

FUNDING SUPPORT

This study was supported by an appointment to the Research Fellowship Program at the Office of New Drugs of the Center for Drug Evaluation and Research (US Food and Drug Administration), which was administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the Food and Drug Administration. This study also involved secondary analysis of data from the Patient Reported Outcomes to Enhance Cancer Treatment (PRO-TECT) trial, which was supported by funding from the Patient-Centered Outcomes Research Institute (award IHS-1511-33392). The PRO-TECT trial was sponsored by Alliance Foundation Trials.

CONFLICT OF INTEREST DISCLOSURES

Christina Mangir reports a payment from Genentech and stock held by an immediate family member in Roche and Bristol-Myers Squibb. Ethan M. Basch reports personal fees from AstraZeneca, Carevive Systems, Navigating Cancer, and Sivan Healthcare. Lynn Howie reports stock ownership at Roche and current employment at Flatiron Health which is an independent subsidiary of Roche. The other authors made no disclosures.

AUTHOR CONTRIBUTIONS

Meena N. Murugappan: Conceptualization, methodology, formal analysis, visualization, writing–original draft, and project administration. **Bellinda L. King-Kallimanis:** Conceptualization, methodology, data curation, writing–review and editing, and supervision. **Christina Mangir:** Writing–review and editing. **Lynn Howie:** Writing–review and editing. **Vishal Bhatnagar:** Writing–review and editing. **Julia A. Beaver:** Writing–review and editing. **Ethan M. Basch:** Writing–review and editing and funding acquisition. **Sydney R. Henson:** Data curation. **Paul G. Kluetz:** Writing–review and editing and funding acquisition.

REFERENCES

1. Patient-Focused Drug Development Glossary. US Food and Drug Administration. Updated June 8, 2018. Accessed February 23, 2021. https://www.fda.gov/drugs/development-approval-process-drugs/patient-focused-drug-development-glossary
2. King-Kallimanis BL, Kanapuru B, Blumenthal GM, Theoret MR, Kluetz PG. Age-related differences in patient-reported outcomes in patients with advanced lung cancer receiving anti–PD-1/PD-L1 therapy. *Semin Oncol*. 2018;45:201-209.
3. Kluetz PG, Slagle A, Papadopoulos EJ, et al. Focusing on core patient-reported outcomes in cancer clinical trials: symptomatic adverse events, physical function, and disease-related symptoms. *Clin Cancer Res*. 2016;22:1553-1558.
4. The Voice of the Patient—Breast Cancer. US Food and Drug Administration. Update May 2016. Accessed March 22, 2021. https://fda.report/media/93924/The-Voice-of-the-Patient-Report--Breast-Cancer.pdf
5. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85:365-376.
6. Luckett T, King MT, Butow PN, et al. Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: issues, evidence and recommendations. *Ann Oncol*. 2011;22:2179-2190.
7. Fayers P, Aaronson NK, Bjordal K, Sullivan M. EORTC QLQ–C30 Scoring Manual. European Organisation for Research and Treatment of Cancer; 1995.
8. Montazeri A. Health-related quality of life in breast cancer patients: a bibliographic review of the literature from 1974 to 2007. *J Exp Clin Cancer Res*. 2008;27:32.
9. Demirci S, Eser E, Ozsaran Z, et al. Validation of the Turkish versions of EORTC QLQ-C30 and BR23 modules in breast cancer patients. *Asian Pac J Cancer Prev*. 2011;12:1283-1287.
10. Chen Q, Li S, Wang M, Liu L, Chen G. Health-related quality of life among women breast cancer patients in eastern China. *Biomed Res Int*. 2018;2018:1452635.
11. Alawadhi SA, Ohaeri JU. Validity and reliability of the European Organization for Research and Treatment in Cancer Quality of Life Questionnaire (EORTC QLQ): experience from Kuwait using a sample of women with breast cancer. *Ann Saudi Med*. 2010;30:390-396.
12. Nejjari C, El Fakir S, Bendahhou K, et al. Translation and validation of European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-C30 into Moroccan version for cancer patients in Morocco. *BMC Res Notes*. 2014;7:228.
13. Garin O. Floor effect. In: Michalos AC, ed. Encyclopedia of Quality of Life and Well-Being Research. Springer; 2014:2300.
14. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4:293-307.
15. Cramer D, Howitt D. The Sage Dictionary of Statistics: A Practical Resource for Students in the Social Sciences. Sage; 2004.
16. Garin O. Ceiling effect. In: Michalos AC, ed. Encyclopedia of Quality of Life and Well-Being Research. Springer; 2014:631-633.
17. Bruce B, Fries J, Lingala B, Hussain YN, Krishnan E. Development and assessment of floor and ceiling items for the PROMIS physical function item bank. *Arthritis Res Ther*. 2013;15:R144.
18. Piccinin C, Kuliś D, Bottomley A, Martinelli F, Pe M, Groenvold M. From item to interface: updating the EORTC Item Library using results from a qualitative content analysis. Poster presented at: 27th Annual Conference of the International Society for Quality of Life Research; October 21-24, 2020.
19. Patient-Reported Outcomes Measurement Information System. HealthMeasures. Accessed September 27, 2020. https://www.healthmeasures.net/explore-measurement-systems/promis
20. Petersen MA, Aaronson NK, Arraras JI, et al. The EORTC CAT Core—the computer adaptive version of the EORTC QLQ-C30 questionnaire. *Eur J Cancer*. 2018;100:8-16.
21. Petersen MA, Aaronson NK, Arraras JI, et al. The EORTC computer-adaptive tests measuring physical functioning and fatigue exhibited

high levels of measurement precision and efficiency. *J Clin Epidemiol*. 2013;66:330-339.
22. Fries JF, Witter J, Rose M, Cella D, Khanna D, Morgan-DeWitt E. Item response theory, computerized adaptive testing, and PROMIS: assessment of physical function. *J Rheumatol*. 2014;41: 153-158.
23. Bass M, Morris S, Neapolitan R. Utilizing multidimensional computer adaptive testing to mitigate burden with patient reported outcomes. *AMIA Annu Symp Proc*. 2015;2015:320-328.
24. Gausden EB, Levack A, Nwachukwu BU, Sin D, Wellman DS, Lorich DG. Computerized adaptive testing for patient reported outcomes in ankle fracture surgery. *Foot Ankle Int*. 2018;39:1192-1198.
25. Bennett AV, Reeve BB, Basch EM, et al. Evaluation of pedometry as a patient-centered outcome in patients undergoing hematopoietic cell transplant (HCT): a comparison of pedometry and patient reports of symptoms, health, and quality of life. *Qual Life Res*. 2016;25:535-546.
26. Mustian KM, Sprod LK, Janelsins M, Peppone LJ, Mohile S. Exercise recommendations for cancer-related fatigue, cognitive impairment, sleep problems, depression, pain, anxiety, and physical dysfunction: a review. *Oncol Hematol Rev*. 2012;8:81-88.
27. McTiernan A, Friedenreich CM, Katzmarzyk PT, et al. Physical activity in cancer prevention and survival: a systematic review. *Med Sci Sports Exerc*. 2019;51:1252-1261.
28. Spei ME, Samoli E, Bravi F, La Vecchia C, Bamia C, Benetou V. Physical activity in breast cancer survivors: a systematic review and meta-analysis on overall and breast cancer survival. *Breast*. 2019;44:144-152.