statistics in oncology

Toward Incorporating Health-Related Quality of Life as Coprimary End Points in Clinical Trials: Time to Achieve Clinical Important Differences and QoL Profiles

Val Gebski, MStat[1]; Andreas Obermair, MD[2]; and Monika Janda, PhD[3]

abstract

PURPOSE Besides morbidity and mortality, quality of life (QoL) is a key outcome of cancer treatments. Trials on the basis of clinical outcomes have expectations that QoL outcomes can be either tolerated or improved. Simultaneously considering QoL and clinical outcomes is challenging with lack of suitable metrics allowing incorporation of QoL as coprimary end points in clinical trial design and utilization of hierarchical hypothesis testing.

METHOD We propose combining time to achieving a minimal clinically important difference (MCID) and probabilities of a MCID occurring in each QoL domain to provide QoL metrics analogous to those used for clinical end points. For QoL domains of interest, these yield QoL profiles, time to MCID, and number needed to treat. Incorporation of QoL as coprimary end points in clinical trial designs through hierarchical hypothesis testing can easily be achieved. The noninferiority designed Laparoscopic Approach to Carcinoma of the Endometrium trial, evaluating laparoscopic versus open abdominal surgery for endometrial cancer with Functional Assessment of Cancer Therapy–General QoL domains, is used to illustrate the usefulness of these metrics.

RESULTS This analysis revealed that laparoscopic surgery had a significant shorter time to MCID for physical and functional well-being QoL domains (physical mean: 1.5 months, 95% CI, 0.5 to 2.6; $P = .002$; and functional mean: 1.4 months; 95% CI, 0.4 to 2.4; $P = .003$) than abdominal surgery, but little difference between the two approaches for psychologic social and emotion well-being. Probability profile plots show a consistent > 2-fold higher chance of attaining a MCID for physical and functional well-being over time for laparoscopic compared with abdominal surgery.

CONCLUSION This analysis reinforces the potential value of novel MCID metrics and their usefulness in raising the profile of QoL outcomes to complement clinical end points. The methods will allow health professionals to counsel patients about QoL outcomes and clinical outcomes simultaneously.

J Clin Oncol 40:2378-2388. © 2022 by American Society of Clinical Oncology

INTRODUCTION

With increasing use of chemotherapy in the treatment of cancer, the publication by Priestman and Baum[1] in mid-1970s heralded a new wave of research on the impact of toxic treatments on patients' well-being and quality of life (QoL),[2–4] addressing questions of how to best treat the patient and not just the disease. Irrespective of the QoL tool used, descriptive analyses remain the primary method used by researchers to assess QoL outcomes. Typically, plots of average QoL domain scores, together with error bars, are presented,[5–7] or comparisons between interventions within/across QoL domains at different time points are made to contrast QoL outcomes across groups of interest.[7–9] Additionally, methods such as quality-adjusted time without symptoms or toxicity,[10,11] quality-adjusted progression-free and overall survival,[11,12] and patient preferences with respect to choice of therapy[13–15] aim to integrate QoL with treatment outcomes. Other approaches include time to QoL deterioration/benefit where attaining or exceeding minimum clinically important difference (MCID) or threshold of QoL changes from baseline is defined as an event and the time to event is represented by Kaplan-Meier curves, compared using log-rank tests or proportional hazards (PH) modeling.[16–19]

Despite these advances and multiple methods of analyzing and interpreting QoL results, greater use of

ASSOCIATED CONTENT
Data Supplement
Protocol

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on April 12, 2022 and published at ascopubs.org/journal/jco on May 16, 2022: DOI https://doi.org/10.1200/JCO.21.02750

CONTEXT

Key Objective

Despite their importance in clinical decision, primary end points on the basis of health-related quality of life (QoL) are uncommon in many clinical trials. We develop novel metrics which, being analogous to common clinical end points, can be used as end points for health-related QoL questionnaires. These can be incorporated as coprimary end points through hierarchical testing procedures to support clinical findings to enhance the impact of interventions to individual patients and enhance clinical decision making.

Knowledge Generated

Applied to a noninferiority surgical trial (Laparoscopic Approach to Carcinoma of the Endometrium), QoL profiles show consistent major benefit for physical functional well-being domains. The profiles can be used to construct QoL hypotheses as coprimary end points and incorporated into trial designs, acknowledging the importance of impact of interventions on QoL domains.

Relevance

The importance of the impact of interventions on QoL cannot be overstated. Although noninferiority and maintenance trials focus on clinical end points, the addition of QoL end points has been challenging. We have developed easily interpretable QoL metrics when combined with the clinical outcomes, provide a comprehensive description of the effect of interventions both clinically and individual patient's QoL.

targeted and immune therapy in oncology has renewed the interest in studying treatment benefits and toxicity.[20] Toxicity trajectories[21] and toxicity profiles[22] make it easier for clinicians and patients to understand the relative toxicity profile of treatment options with comparable survival outcomes. Toxicity and QoL experiences are interdependent, although toxicity profiles fail to take into account QoL improvements in patients who are symptomatic and respond to treatment with reduction in symptoms and improvement in some/all QoL domains. There is an opportunity to adapt these methods to (1) quantify the likelihood that an individual will experience a MCID in each QoL domain by a particular time, and (2) describe the relative benefit/detriment of receiving one treatment versus another over time for specific QoL domains. These likelihoods can be used to quantify potential benefits/detriments for both individuals as well as patient and treatment groups. Such profiles would be especially useful when designing noninferiority studies; these trials require evidence of benefit in key secondary end points, commonly QoL, to inform health policy and practice change. QoL profiles would also assist in the design of trials using QoL end points as primary/coprimary outcomes and/or guide patient-clinician decision making with respect to the choice of therapy.

METHODS

To demonstrate the application of the novel method, we use the Laparoscopic Approach to Carcinoma of the Endometrium (LACE, ClinicalTrials.gov identifier: NCT00096408) trial as an example.[23] The LACE noninferiority randomized phase III trial enrolled 760 patients between 2005 and 2010 comparing total laparoscopic hysterectomy (TLH) to total abdominal hysterectomy (TAH) for women with stage I endometrial cancer. A group of 361 patients were enrolled in a QoL substudy. The LACE study was conducted in accordance with the Declaration of Helsinki, approved by the Ethics Committee of each participating center, and complied with the provisions of the Good Clinical Practice Guidelines. All patients provided written informed consent.

The primary outcome of the LACE trial was 4.5-year disease-free survival (DFS). This study is illustrative in that, although the main clinical interest is DFS, QoL is of major importance to patients (and clinicians). However, changing clinical practice on the basis of patient-rated QoL alone as a primary outcome is difficult and therefore, clinical trials usually specify primary end points related to traditional outcome measures with QoL relegated to secondary objectives. As QoL typically comprises many domains, it is challenging to summarize different benefits into a single parameter, thereby limiting its utility in routine clinical practice. Although methods aimed at combining QoL results over different domains have been proposed,[24–28] their use in practice is limited, possibly because of the difficulty of generalizing the results to inform clinical decision making in individual patients.

Defining the MCID can be consensus-derived or on the basis of patient-clinician preferences. These thresholds should ideally be absolute or relative values, and thresholds derived from magnitude-based differences or effect sizes should be avoided because of their methodologic immaturity.[29–31] As the LACE trial evaluated curative treatments, most participants would expect to return to their usual tasks in a relatively short period, and even transient gains in QoL would be important to patients.

The LACE trial demonstrated noninferiority of DFS at 4.5 years,[23] and a traditional QoL analysis demonstrated

benefits of TLH compared with TAH over a number of QoL domains.[6] Briefly, QoL was measured using the Functional Assessment of Cancer Therapy–General questionnaire, which covers five metrics in the well-being domains: physical (PWB), social (SWB), functional (FWB), emotional (EWB), and endometrial cancer–specific (EnWB). For the LACE trial, the MCID $\geq$ 5 QoL points ($\geq$ 5%) was deemed appropriate[6] for all domains; however, for the novel method proposed here, the thresholds for each domain are not required to be equal. We define an event as the first time an MCID is observed irrespective of the domain. For each domain, we define E_{PWB} as the event if it was first observed in PWB; E_{SWB}, if first observed in SWB; E_{FWB}, if first observed in FWB; E_{EWB}, if first observed in EWB, and E_{EnWB}, if first observed in EnWB.

(A) Calculation of the Probability of an Event in Each QoL Domain Among Those Who Have Attained a MCID

If we consider the subset of patients achieving a MCID in at least one domain, that is, experiencing one of (E_{Pwb}, E_{Swb}, E_{Fwb}, E_{Ewb}, and E_{fnwb}), then the probability that the first MCID is experienced in a particular domain can be estimated by fitting a polytomous logistic regression[32] with domain events as the outcomes, and treatment, *ln*(time to the event) as the predictors. The time component accounts for the duration until an MCID is reached. This regression model provides odds ratios for the treatment and *ln*(time) effects in each domain from which the fitted probabilities of (E_{Pwb}, E_{Swb}, E_{Fwb}, E_{Ewb}, and E_{fnwb}) can be obtained for each treatment and time combination. These probabilities can be viewed as weights for an individual to experience a MCID in each domain. Polytomous logistic regression procedures are provided in most common statistical packages (STATA, SAS, SPSS, and R). Our analyses were performed in the ACCoRD statistical package.[33]

(B) Calculation of the Hazard of the First MCID Over All Domains Over Time

The hazard of an individual experiencing the first MCID at each event time in any domain can be estimated by the Nelson-Aalen[34] method on the basis of all cases. The hazard ratio (HR) of experiencing an MCID (in any domain) of the intervention (TLH) relative to the control (TAH) is estimated using a Cox PH model. These hazards, combined with probabilities obtained in (A) and the HR, give the cumulative probability (CP) of experiencing a MCID event in each domain and treatment over time.[22] The probabilities increase over time, are bounded between 0 and 1, and can be displayed graphically.

(C) Time to the First MCID in Each Domain for Each Treatment Group and Number Needed to Treat

Subtracting the probabilities in (B) from 1.0 provides curves for the time to MCID in each domain (analogous to, but different from, Kaplan-Meier curves). The average time to MCID can be estimated as the area under these curves for each treatment and domain combination. Differences in these areas represent the average additional time taken to experience a MCID between treatment groups. Comparisons of these areas require their variances, which can be estimated using the bootstrap approach.[35] Approximate 95% CIs and associated *P* values for this difference can be obtained using the *t*-test. For a time point of interest, the number needed to treat (NNT)[36] can be obtained as the reciprocal of the difference between these probabilities obtained in (B) for the two treatment groups. The NNT provides the number of patients required to be treated with the favorable treatment (and higher probability of achieving a MCID) to achieve one additional patient attaining a MCID up to this time. A worked example in applying these methods is given in the Data Supplement (online only).

MCID and Competing Clinical Events

The question arises as to how can the occurrence of competing clinical events (death, disease progression, etc) before observing an MCID be accommodated by this approach. If QoL is affected by the disease state (progression and death) preceding any MCID, then deaths/clinical outcomes can be included as an additional category in the polytomous regression. Polytomous regression does not distinguish as to which events are being considered, only that they be mutually exclusive. This gives rise to the category death before MCID and will adjust the estimated QoL probabilities from the model in (A), as well as being propagated through to the hazard calculations. The hazard calculations would include death/clinical outcome before MCID as an event. As survival is usually modeled separately (Kaplan-Meier/PH, etc), the probabilities from the category of death/clinical outcome would be of secondary interest and serve as an adjustment for QoL probability profiles in each domain.

If the number of deaths/clinical events is small, from a pragmatic perspective, they could be considered similarly to loss to follow-up where in both cases further QoL is no longer available. Although not ideal, in practice, any biases induced by this definition should be small and would be confined to the polytomous regression component which, as the model assumptions are minimal, would be robust to issues of the small number of competing risks.

Levels of noncompletion/noncompliance of QoL components in clinical studies have a tendency to be notoriously high. This can be due to disease deterioration/improvement or patient-related factors (eg, age, cultural, and lifestyle) for which the addition of an extra multinomial category may not be sufficient. Assessment of missingness is key, and the impact of QoL missingness (and while may be alleviated through the use of appropriate surrogates[37]) should accompany the proposed method.

RESULTS

Application to the LACE Trial

For this analysis of the LACE trial, Functional Assessment of Cancer Therapy–General results were available for 313 of 361 patients, 128 allocated to TAH and 185 to TLH. All 313

patients completed six visits at baseline, and assessment times varying around weeks 1, 4, 12, and 26 (Fig 1). Generally, adherence to planned protocol (online only) visit times is not strict but will occur in a window around these times. Within a time window, QoL collection broadly coincided with clinic visits or availability of patients from remote centers to attend centrally. The choice of the time window was pragmatic and unrelated to QoL status, minimizing potential bias. Eleven patients (4%) did not achieve an MCID in any domain. Distribution of patients achieving a first MCID was 50 (16%) for PWB, 108 (35%) SWB, 21 (7%) FWB, 33 (32%) EWB, and 20 (6%) EnWB. Patient characteristics are detailed in the study by Janda et al.[6]

The HR of a MCID for TLH:TAH is 1.20 (95% CI, 0.95 to 1.52; $P = .123$), indicating a 20% increased benefit for TLH. The HR is an estimate of the relative effect of the intervention, the purpose being to attenuate the probability, for the intervention, of achieving an MCID in each domain. As the purpose is to identify differences in probabilities within the domains, the requirement for statistical significance in the time to MCID over all domains would be unduly restrictive, and this requirement can be relaxed.

Table 1 shows the monthly cumulative probabilities of attaining a MCID after surgery and QoL domain.

The physical and functional domains demonstrate a higher chance of attaining a MCID for TLH, but this is attenuated for SWB and EnBW at early time points (< 4 months). For each domain, the last row gives the NNT. Five additional patients treated with TLH will yield one additional patient attaining a MCID by 6 months for PWB and FWB.

Probability curves for each domain are displayed in Figure 2, showing the trajectory of the chance MCID.

QoL profile plots for each domain can be obtained as the ratio of the probabilities in Figure 2. The ratio of the CP over time of TLH to TAH is shown in Figure 3, for all the domains.

For physical domains, patients receiving TLH have an increased chance of experiencing a MCID in PWB and FWB > 2-fold within 3 months, and this reduced to 1.6-fold after 3 months for FWB. The nonphysical domains, SWB and EWB, show little separation between the two treatments with TAH initially having a slightly higher chance of

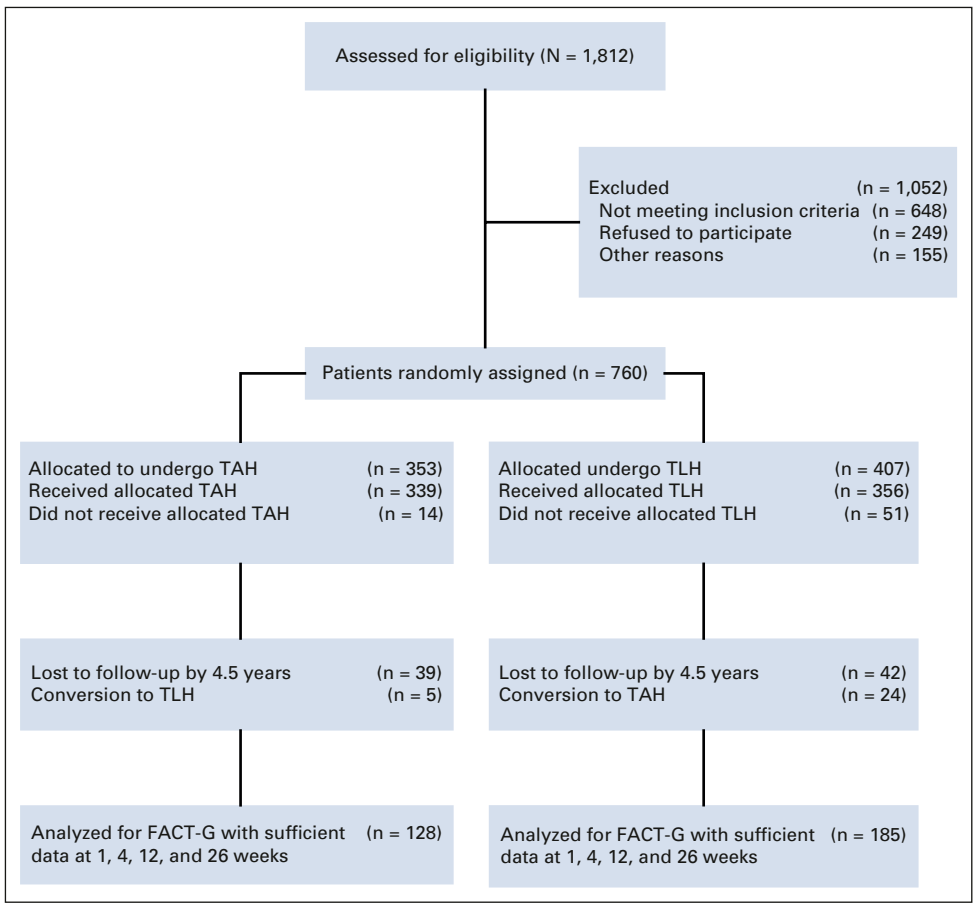

FIG 1. CONSORT diagram. FACT-G, Functional Assessment of Cancer Therapy–General; TAH, total abdominal hysterectomy; TLH, total laparoscopic hysterectomy.

TABLE 1. Quality-of-Life Domain Where the First Minimal Clinically Important Difference (≥ 5 points) Is Experienced: Cumulative Probabilities of the Event by Treatment Over the First 6 Months After Surgery

		Months Since Random Assignment					
Domain	**Treatment**	**1**	**2**	**3**	**4**	**5**	**6**
PWB	TAH	0.112	0.198	0.210	0.284	0.284	0.284
	TLH	0.242	0.390	0.411	0.449	0.462	0.478
NNT		8	5	5	6	6	5
SWB	TAH	0.356	0.497	0.513	0.592	0.592	0.592
	TLH	0.399	0.540	0.557	0.583	0.591	0.600
NNT		23	23	23	a	a	a
FWB	TAH	0.030	0.076	0.085	0.149	0.149	0.149
	TLH	0.096	0.213	0.236	0.286	0.307	0.335
NNT		15	8	7	7	6	5
EWB	TAH	0.367	0.513	0.530	0.611	0.611	0.611
	TLH	0.363	0.500	0.517	0.543	0.551	0.560
NNT		a	a	a	a	a	a
EnWB	TAH	0.056	0.120	0.131	0.205	0.205	0.205
	TLH	0.080	0.160	0.175	0.204	0.216	0.231
NNT		42	25	23	a	91	38

Abbreviations: EnWB, endometrial cancer–specific well-being; EWB, emotional well-being; FWB, functional well-being; NNT, number of patients needed to be treated with TLH in order for one extra patient to achieve a minimal clinically important difference over that seen in TAH at different study follow-up times; PWB, physical well-being; SWB, social well-being; TAH, total abdominal hysterectomy; TLH, total laparoscopic hysterectomy.

[a]No advantage of TLH over TAH for this domain at this follow-up time point.

MCID for EnWB (approximately 1.3-fold increase), which then diminishes.

These profiles serve to highlight the trajectory of patient's QoL experiences over time in both absolute and relative terms. The cumulative probabilities in Figure 2, when subtracted from 1.0, give the time to MCID. The area under this curve gives the average time to MCID in each QoL domain and treatment group. The average time to MCID is given in Table 2, together with the difference, 95% CIs, and *P* value.

Over the 6-month LACE trial follow-up period, PWB and FWB showed a statistically significant shorter average time to MCID of approximately 1.5 months with only a modest nonsignificant shorter time of approximately 0.5 months for the other three QoL domains. Figure 4 presents a graphical display of the magnitude of the time to MCID with the shaded portion representing the time gain of TLH over TAH, whereas the cross-hatched areas (for EWB) show time deficit.

These results enhance and extend the conclusions in the original report,[6] with the areas in Figure 4 clearly showing the pattern of benefit: patients who received a TLH achieved a sustained large improvement in the time to MCID for PWB and FWB, and a small improvement for SWB and EnWB, while there was a small prolongation of time to MCID with TAH for EWB.

Application of the Novel Methodology for Clinical Trial Design

Our metrics, building on existing QoL scales and closely aligning with patient experiences, can be incorporated as coprimary outcomes into future clinical trial designs to complement their clinical objectives. The simplest approach is to order the hypotheses according to their importance. The clinical hypothesis (H_1) would be considered of primary interest. Benefits in QoL domains (H_2) are also of key interest but would generally require rejection of the null hypothesis defined in H_1. The principle underpinning this approach is that QoL questions will *only be compared* if H_1 is statistically significant (benefit/noninferiority demonstrated). The level of significance α, used in the H_1 comparison, is the same if we proceed to perform QoL comparisons (H_2) subject to the above principle. Thus, if three QoL domains are to be compared in H_2, the *overall* significance level will be at α, although multiple comparison adjustments are required if more than one domain is to be compared. More complex multiple testing schemes can also be considered,[38] say, further levels of hypotheses (H_3, H_4…) comprising less important domains to those in H_2. Such schemes provide flexibility in assigning significance levels to the different domains without requiring all the significance levels to be the same, and stronger levels of evidence (smaller α) may be assigned to domains where larger benefits are necessary to support the intervention as

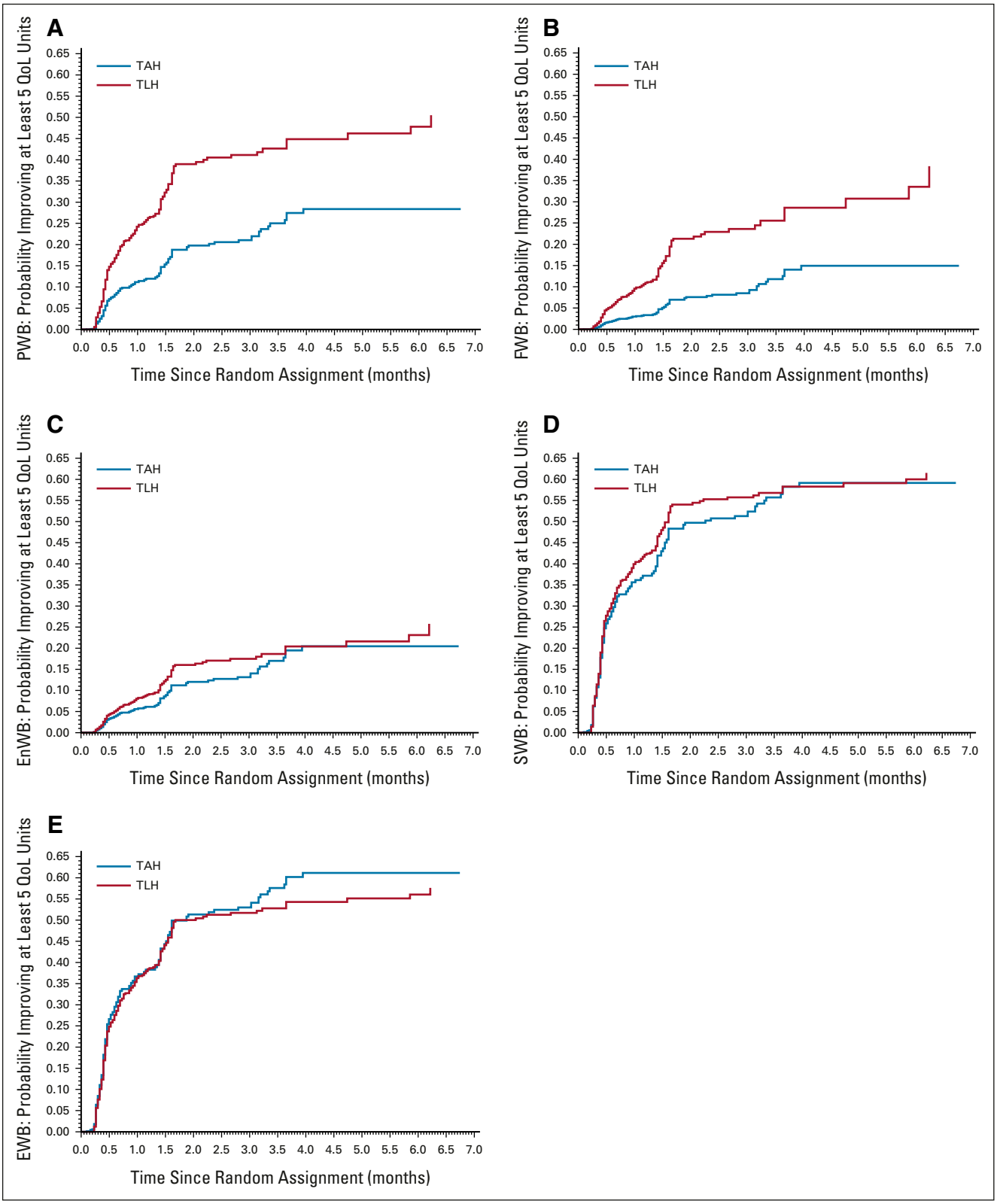

FIG 2. Cumulative probability QoL profiles of attaining a minimal clinically important difference at or before a time of interest: (A) PWB, (B) FWB, (C) EnWB, (D) SWB, and (E) EWB. EnWB, endometrial cancer–specific well-being; EWB, emotional well-being; FWB, functional well-being; PWB, physical well-being; QoL, quality of life; SWB, social well-being; TAH, total abdominal hysterectomy; TLH, total laparoscopic hysterectomy.

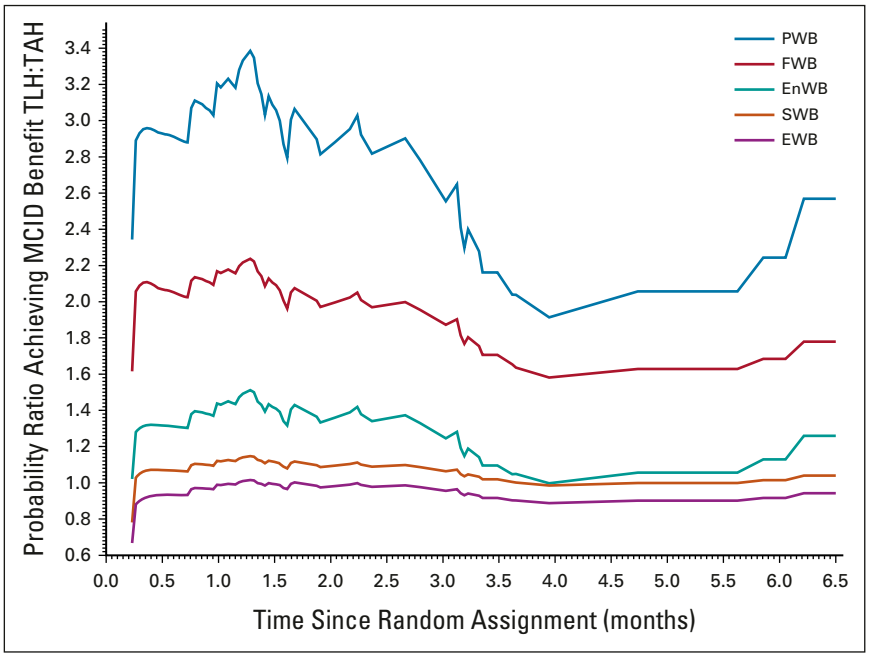

FIG 3. Probability ratio profiles: MCID of TLH relative to TAH over time for all five quality-of-life domains. EnWB, endometrial cancer–specific well-being; EWB, emotional well-being; FWB, functional well-being; MCID, minimal clinically important difference; PWB, physical well-being; SWB, social well-being; TAH, total abdominal hysterectomy; TLH, total laparoscopic hysterectomy.

desirable. If the domains in H_2 have a moderate to high correlation, a gain in each of the significance levels is obtained while still retaining an overall significance level of α. For example, if H_2 comprised only PWB and FWB and they were uncorrelated, each would be compared at significance level 2.5% ($\alpha/2$) for an overall $\alpha = 5\%$. By contrast, a strong correlation between PWB and FWB of 0.7 allows for each for comparison at a significance level of 2.9% while retaining an overall significance level of 5%.[39] In the LACE trial, the correlation between FWB and EWB was 0.68. For interventions with time-to-event outcomes, measures of benefit are either hazard or odds ratios or differences in proportions at a fixed time point. The QoL metric of time to MCID dovetails into these definitions with an interpretation easily understood by patients and clinicians. These calculations are not intended to be interpreted as poststudy power (which is scientifically flawed) but serve to illustrate how QoL end points can be incorporated into designs as coprimary end points, and potential sample size calculations may be performed to account for multiple comparisons and correlations among end points.

Where H_1 is aiming to show superiority, when interventions are associated with increased toxicity, the MCID is the time to a deterioration. Longer time to MCID is desirable, and H_2 would be formulated as a noninferiority hypothesis: the time to deterioration would be no shorter than x months. Precisely which QoL domains would fall into H_2 requires some thought as, with adjustment for multiple comparisons, noninferiority could have an impact on the study sample size. To illustrate, the second-generation epidermal growth factor receptor tyrosine kinase inhibitor therapy (afatinib) for non–small-cell lung cancer demonstrated a progression-free survival advantage while also showing higher toxicity rates compared with the first-generation therapies (gefitinib).[20] The QoL component of interest is related to skin toxicity and measured by the Skindex-16[40] SRQOL. A noninferiority margin for deterioration of 0.75 months, standard deviation of 2 months, 90% power, and one-sided $\alpha = 5\%$ yields 246 patients for the H_2 comparison. The sample sizes in these second-generation studies were in excess of 246.

A second potential end point metric is the difference between the probabilities of achieving an MCID at a time point of interest between the treatment groups. This is directly related to the NNT with design considerations and sample size calculations following standard methods on the basis of differences between two proportions.[41] In these cases, we require the probability of achieving a MCID at the fixed time in the control group, which would come from a pilot study.

DISCUSSION

Strengths of this novel methodology approach presented in this paper include (1) relevant QoL domains are considered simultaneously; (2) MCIDs are incorporated in a time-to-event framework yielding CP estimates together with the NNT, and (3) MCID can be co-considered together with clinical outcomes, thus allowing for considering trade-offs

TABLE 2. Average Time to Attaining a MCID in Each Quality-of-Life Domain, Differences and 95% CI

Domain	Treatment	Time to MCID	Difference (TAH-TLH)	95% CI	*P*
PWB	TAH	5.2	1.5	0.5 to 2.6	.002
	TLH	3.7			
SWB	TAH	3.3	0.5	–0.5 to 1.4	.163
	TLH	2.8			
FWB	TAH	6.0	1.4	0.4 to 2.4	.003
	TLH	4.6			
EWB	TAH	3.2	0.1	–0.8 to 1.0	.383
	TLH	3.1			
EnWB	TAH	5.7	0.7	–0.3 to 1.7	.086
	TLH	5.0			

Abbreviations: EnWB, endometrial cancer–specific well-being; EWB, emotional well-being; FWB, functional well-being; MCID, minimal clinically important difference; PWB, physical well-being; SWB, social well-being; TAH, total abdominal hysterectomy; TLH, total laparoscopic hysterectomy.

in superiority trials. Ratios of these estimates produce the trajectory over time of QoL benefit/deterioration identifying time frames at which QoL changes are observed and the relative magnitude of such changes. Estimates can also be displayed as time-to-MCID curves allowing calculation of the average time to MCID for each domain. These can be formally compared by treatment to ascertain the extent to which QoL changes over the study period. These principles are illustrated for the LACE trial which, being a noninferiority surgical curative trial, demonstrated that QoL benefits were key trial results to influence clinical practice. It is not the anticipated behavior of the new hazard estimate to be overly sensitive to low event rates that may occur in a particular QoL domain. In the event of extremely small event rates, excluding the problematic domain(s) in the basket of domains being considered may be a more sensible (because of numerical instability) strategy, and these domain(s) may be described separately. This would be an issue with most statistical methods and not just confined to this one.

For superiority studies where time to QoL deterioration is a key outcome, the methods and interpretation proposed here are analogous as for QoL benefits. Profile plots, average time to deterioration, and NNT (to avoid one additional MCID) would be the metric. In palliative settings, long-term treatment is associated with gradual/eventual deterioration in QoL. Here, the event of interest is deterioration-free survival, and MCID thresholds are minimal amounts of acceptable QoL detriment. Deterioration thresholds of ≤ 10 QoL points have been reported in oncology studies of immunotherapy.[42] For QoL deterioration, the methods and interpretation proposed here are analogous to QoL benefits. Where substantial uncertainty as to the threshold for a MCID in each of the QoL domains, pilot studies on the basis of time trade-offs between clinical benefit/noninferiority and QoL detriment/gains[43] can be conducted to estimate the intersection between QoL and clinical gains yielding the MCID for each domain. Clinically, the novel MCID methodology can be used to counsel patients about trading survival benefits against QoL impairment.

We have used a noninferiority design in a surgical trial (with curative intent) in oncology to illustrate the proposed concepts. The QoL domains (whether from the Short Form-36 questionnaire or the Quality of Life Questionnaire C30) provide generic domains across many disease conditions. In setting up the hierarchy of QoL hypotheses, for curative interventions, physical/functions domains may be of key importance. For noncurative conditions (eg, hemodialysis), the focus of QoL focus may be on the emotional/social improvement, and these domains would feature higher in the hypothesis hierarchy. The requirement that all QoL domains demonstrate benefit/nondeterioration would not be common.

It is instructive to contrast the cumulative incidence (CuI) curves proposed by Gray in the context of competing risk analysis[44] to the proposed CP curves. The CuI curves, related to the Kaplan-Meier survival curve S(t), partitions S(t) such that $1 - S(t) = \sum$[CuI for each event type]. All event types (the QoL domains) are considered jointly so that, at any time point, the sum of the CuI probabilities over all event types is ≤ 1. For a particular event type, the CuI curve only increases at the time points when events of that type occur (step-function) and is constant between events. Comparisons of the CuI curves using statistical tests (analogous to the log-rank test) have been developed using these definitions. However, when there are long gaps between events, combined with a reduced number of patients at risk, interpretation of the cumulative probabilities in a clinical setting can be challenging. While this pattern is illustrated in the context of the Kaplan-Meier curve through the TIELCAP report,[45] CuI curves would exhibit similar patterns, and such patterns are not uncommon when analyzing local relapse in surgical or radiation therapy studies requiring the use of competing risk methods.

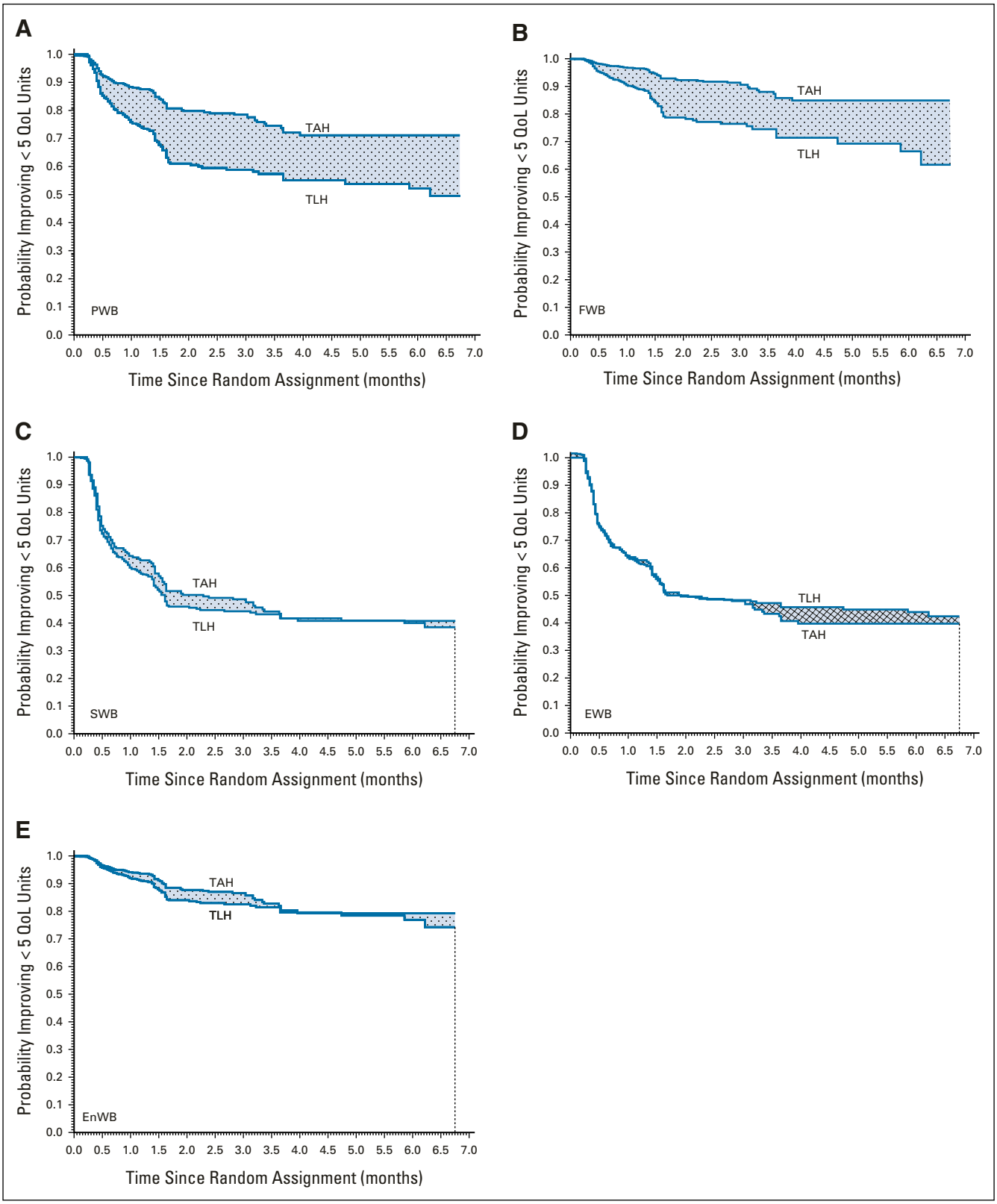

FIG 4. Difference in time to MCID (shaded) for TAH and TLH: (A) PWB, (B) FWB, (C) SWB, (D) EWB, and (E) EnWB. Time to MCID favors TAH in (D). Shaded area represents the difference in the reduction in time to MCID for TLH compared with TAH with the exception of (D) where the direction is reversed. EnWB, endometrial cancer–specific well-being; EWB, emotional well-being; FWB, functional well-being; MCID, minimal clinically important difference; PWB, physical well-being; QoL, quality of life; SWB, social well-being; TAH, total abdominal hysterectomy; TLH, total laparoscopic hysterectomy.

CP curves, by contrast, having modeled joint event probabilities (which sum to 1) and a time component, adjust the risk of an event at each event time (irrespective of event type) and, for each event type, increase less abruptly than the CuI. For each event type, the CP curve contains more probability components at each time point than the CuI without the constraint that the sum of the probabilities over each event type be ≤ 1. These can be viewed as being marginal cumulative probabilities, yielding separate probabilities of each event type over time separately. Comparisons of the CP curves, as either proportions at specified time points, or areas under the curve, require computational techniques such as the bootstrap.

The new methods proposed here also overcome some of the sample size issues in noninferiority trials where multiple end points are considered. We have outlined how QoL can be incorporated in primary study design as it is of key importance, particularly to patients and supportive care professionals.

Our novel methods for time to the first MCID can be extended to studies where outcomes give rise to recurrent episodes of achieving a MCID event. In long-term QoL follow-up studies, subjects may experience repeat MCID events over time because of their disease history or ongoing treatment. In these cases, each event recurrence is considered separately. Thus, if considering the third MCID event, the polytomous model is fitted to just the cohort who have experienced the third MCID and the probabilities for each domain estimated. The hazard is calculated using just those patients who have experienced any event up until the third MCID (patients experiencing 0, 1, or 2 MCIDs are censored at the time of their last follow-up). The HR for the third event is estimated from PH recurrent event methods.[22,46] The analysis is analogous to that which is used for the first event.

Analogous methods have also been developed to examine, within a QoL domain, different (ordered) thresholds,[46] for example, small, moderate, or large improvement/deterioration, and can yield estimates of probabilities that an individual will achieve on each of these categories over time.

In conclusion, novel metrics for QoL outcomes can assist in interpreting the impact of new interventions, medical decision making, and translating benefits/detriments for individual patients. Combining with the chance of a MCID with the probability that a MCID will occur in each QoL domain provides a framework for incorporating QoL outcomes with traditional clinical end points. Embedding these metrics into clinical trial designs provides an additional dimension to evaluating QoL benefits as primary questions.

AFFILIATIONS

[1]NHMRC Clinical Trials Centre, University of Sydney, Sydney, New South Wales, Australia
[2]Queensland Centre for Gynaecological Cancer Research, University of Queensland, Brisbane, Queensland, Australia
[3]Centre for Health Services Research, The University of Queensland, Brisbane, Queensland, Australia

CORRESPONDING AUTHOR

Val Gebski, MStat, NHMRC Clinical Trials Centre, Locked Bag 77, Camperdown 1450, New South Wales, Australia; e-mail: val.gebski@sydney.edu.au.

SUPPORT

Supported by the National Health and Medical Research Council Program Grant No. 1150467 awarded to the NHMRC Clinical Trials Center, University of Sydney.

CLINICAL TRIAL INFORMATION

NCT00096408

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at DOI https://doi.org/10.1200/JCO.21.02750.

AUTHOR CONTRIBUTIONS

Conception and design: All authors
Administrative support: Andreas Obermair
Provision of study materials or patients: Andreas Obermair, Monika Janda
Collection and assembly of data: All authors
Data analysis and interpretation: Val Gebski
Manuscript writing: All authors
Final approval of manuscript: All authors
Accountable for all aspects of the work: All authors

REFERENCES

1. Priestman T, Baum M: Evaluation of quality of life in patients receiving treatment for advanced breast cancer. Lancet 307:899-901, 1976
2. Coates A, Abraham S, Kaye S, et al: On the receiving end—Patient perception of the side-effects of cancer chemotherapy. Eur J Clin Oncol 19:203-208, 1983
3. Coates A, Dillenbeck C, McNeil D, et al: On the receiving end-II. Linear analogue selfassessment (LASA) in evaluation of aspects of the quality of life of cancer patients receiving therapy. Eur J Cancer Clin Oncol 19:1633-1637, 1983
4. Coates A, Glasziou P, McNeil D: On the receiving end—III measurement of quality of life during cancer chemotherapy. Ann Oncol 1:213-217, 1990
5. Brundage M, Gropp M, Mefti F, et al: Health related quality of life in recurrent platinum-sensitive ovarian cancer—Results rrom the CALYPSO trial. Ann Oncol 23:2020-2027, 2012
6. Janda M, Gebski V, Brand A, et al: Quality of life after total laparoscopic hysterectomy versus total abdominal hysterectomy for stage I endometrial cancer (LACE): A randomised trial. Lancet Oncol 11:772-780, 2010
7. Sanda M, Dunn R, Michalski J, et al: Quality of life and satisfaction with outcome among prostate-cancer survivors. N Engl J Med 358:1250-1261, 2008

8. Kornblith A, Huang H, Walker J, et al: Quality of life of patients with endometrial cancer undergoing laparoscopic international federation of gynecology and obstetrics staging compared with laparotomy: A gynecologic oncology group study. J Clin Oncol 27:5337-5342, 2009
9. Mark D, Pan W, Clapp-Channing N, et al: Quality of life after late invasive therapy for occluded arteries. N Engl J Med 360:774-783, 2009
10. Friedlander M, Gebski V, Gibbs E, et al: Health-related quality of life and patient-centred outcomes support the clinical benefit of prolongation of progression-free survival with olaparib maintenance following chemotherapy in patients with germline BRCA-mutated platinum-sensitive relapsed serous ovarian cancer: SOLO2 randomised phase 3 trial. Lancet Oncol 19:1126-1134, 2018
11. Lenderking W, Gelber R, Cotton D, et al: Evaluation of the quality of life associated with zidovudine treatment in asymptomatic human immunodeficiency virus infection. N Engl J Med 330:738-743, 1994
12. Verduyn S, Biesma B, Schramel F, et al: Estimating quality adjusted progression free survival of first-line treatments for EGFR mutation positive non small cell lung cancer patients in The Netherlands. Health Qual Life Outcomes 10:108, 2012
13. Duric V, Stockler M, Heritier S, et al: Patients' preferences for adjuvant chemotherapy in early breast cancer: What makes AC and CMF worthwhile now? Ann Oncol 16:1786-1794, 2006
14. Simes R, Coates A: Patient preferences for adjuvant chemotherapy of early breast cancer: How much benefit is needed? J Natl Cancer Inst Monogr 30:146-152, 2001
15. Stuart-Harris R, Simes R, Coates A, et al: Patient treatment preference in advanced breast cancer: A randomized cross-over study of doxorubicin and mitozantrone. Eur J Clin Oncol 23: 557-561, 1987
16. Anota A, Hamidou Z, Paget-Bailly S, et al: Time to health-related quality of life score deterioration as a modality of longitudinal analysis for health-related quality of life studies in oncology: Do we need RECIST for quality of life to achieve standardization? Qual Life Res 24:5-18, 2013
17. Bonnetain F, Dahan L, Maillard E, et al: Time until definitive quality of life score deterioration as a means of longitudinal analysis for treatment trials in patients with metastatic pancreatic adenocarcinoma. Eur J Cancer 46:2753-2762, 2010
18. Brahmer J, Rodríguez-Abreu D, Robinson A, et al: Health-related quality-of-life results for pembrolizumab versus chemotherapy in advanced, PD-L1-positive NSCLC (KEYNOTE-024): A multicentre, international, randomised, open-label phase 3 trial. Lancet Oncol 18:1600-1609, 2017
19. Hamidou Z, Dabakuyo T, Mercier M, et al: Time to deterioration in quality of life score as a modality of longitudinal analysis in patients with breast cancer. Oncologist 16:1458-1468, 2011
20. Ding P, Lord S, Val Gebski V, et al: Risk of treatment-related toxicities from EGFR tyrosine kinase inhibitors: A meta-analysis of clinical trials of gefitinib, erlotinib, and afatinib in advanced EGFR-mutated non–small cell lung cancer. J Thorac Oncol 12:633-643, 2017
21. Francis K, Gebski V, Lord S, et al: Reporting the trajectories of adverse events over the entire treatment course in patients with recurrent platinum-sensitive ovarian cancer treated with platinum based combination chemotherapy regimens: A graphical approach to trial adverse event reporting. Eur J Cancer 148:251-259, 2021
22. Gebski V, Marschner I, Asher R, et al: Using recurrent time-to-event models with multinomial outcomes to generate toxicity profiles. Pharm Stat 20:840-849, 2021
23. Janda M, Gebski V, Davies L, et al: Effect of total laparoscopic hysterectomy vs total abdominal hysterectomy on disease-free survival among women with stage I endometrial cancer A randomized clinical trial. J Am Med Assoc 317:1224-1233, 2017
24. O'Brien P: Comparing samples with multiple endpoints. Biometrics 40:1079-1087, 1984
25. Olschewski M, Schumacher M: Statistical analysis of quality of life data in cancer clinical trials. Stat Med 9:749-763, 1990
26. Follmann D, Wittes J, Cutler J: The use of subjective rankings in clinical trials with an application to cardiovascular disease. Stat Med 11:427-437, 1992
27. Testa M, Nackley J: Methods for quality-of-life studies. Annu Rev Public Health 15:535-559, 1994
28. Lumley T, Simes R, Gebski V, et al: Combining components of quality of life to increase precision and evaluate trade-offs. Stat Med 20:3231-3249, 2001
29. Welsh AH, Knight EJ: "Magnitude-based inference": A statistical review. Med Sci Sports Exerc 47:874-884, 2015
30. Nevill A, Williams A, Boreham C, et al: Can we trust "Magnitude-based inference"? J Sports Sci 36:2769-2770, 2018
31. Sainani K: The problem with "Magnitude-based inference". Med Sci Sports Exerc 50:2166-2176, 2018
32. Dobson AJ: An Introduction to Generalized Linear Models (ed 2). Boca Raton, FL, Chapman & Hall/CRC, 2002
33. ACCoRD (Analysis of Censored and Correlated Data). Eastwood, NSW Australia, Boffin Software, 2018
34. Klein JP, Moeschberger ML: Survival Analysis Techniques for Censored and Truncated Data (ed 2). New York, NY, Springer, 2003
35. Walters S, Campbell M: The use of bootstrap methods for analysing health-related quality of life outcomes (particularly the SF-36). Health Qual Life Outcomes 2:70, 2004
36. Altman D, Andersen P: Calculating the number needed to treat for trials where the outcome is time to an event. Br Med J 319:1492-1495, 1999
37. Beller E, Tattersall M, Lumley T, et al: Improved quality of life with megestrol acetate in patients with endocrine-insensitive advanced cancer: A randomised placebo-controlled trial. Ann Oncol 8:277-283, 1997
38. Wang D, Li Y, Wang X, et al: Overview of multiple testing methodology and recent development in clinical trials. Contemp Clin Trials 45:13-15, 2015
39. Gupta S: Probability integrals of multivariate normal and multivariate t. Ann Math Stat 34:792-828, 1963
40. Tseng L, Chen K, Wang C, et al: Effects of tyrosine kinase inhibitor therapy on skin toxicity and skin-related quality of life in patients with lung cancer—An observational study. Medicine (Baltimore) 99:e20510, 2020
41. Bell M, Teixeira-Pinto A, McKenzie J, et al: A myriad of methods: Calculated sample size for two proportions was dependent on the choice of sample size formula and software. J Clin Epidemiol 67:601-605, 2014
42. Vaughn D, Bellmunt J, Fradet Y, et al: Health-related quality-of-life analysis from KEYNOTE-045: A phase III study of pembrolizumab versus chemotherapy for previously treated advanced urothelial cancer. J Clin Oncol 36:1579-1587, 2018
43. Brown A, Atyeo J, Field N, et al: Evaluation of patient preferences towards treatment during extended hours for patients receiving radiation therapy for the treatment of cancer: A time trade-off study. Radiother Oncol 90:247-252, 2009
44. Tai BC, Machin D, White I, et al: Competing risks analysis of patients with osteosarcoma: A comparison of four different approaches. Stat Med 20:661-684, 2001
45. International Early Lung Cancer Action Program Investigators, Henschke CI, Yankelevitz DF, et al: Survival of patients with stage I lung cancer detected on CT screening. N Engl J Med 355:1763-1771, 2006
46. Gebski V, Byth K, Asher R, et al: Recurrent time-to-event models with ordinal outcomes. Pharm Stat 20:77-92, 2021

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Toward Incorporating Health-Related Quality of Life as Coprimary End Points in Clinical Trials: Time to Achieve Clinical Important Differences and QoL Profiles

The following represents disclosure information provided by authors of this manuscript. All relationships are considered compensated unless otherwise noted. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/authors/author-center.

Open Payments is a public database containing information reported by companies about payments made to US-licensed physicians (Open Payments).

Andreas Obermair
Stock and Other Ownership Interests: SurgicalPerformance Pty Ltd
Honoraria: Baxter
Consulting or Advisory Role: OR Company
Research Funding: Medtronic
Patents, Royalties, Other Intellectual Property: I own shares in SurgicalPerformance Pty Ltd

Val Gebski

This author is a member of the *Journal of Clinical Oncology* Editorial Board. Journal policy recused the author from having any role in the peer review of this manuscript.

Monika Janda
Stock and Other Ownership Interests: SurgicalPerformance Pty Ltd (I)
Honoraria: Baxter (I)
Consulting or Advisory Role: OR company (I)
Research Funding: Medtronic
Patents, Royalties, Other Intellectual Property: Trademark for SurgicalPerformance (I)

No other potential conflicts of interest were reported.