Analyse longitudinale des données de qualité de vie relative à la santé en cancérologie : vers une standardisation de la méthode du temps jusqu'à détérioration

Emilie Charton

▶ **To cite this version:**

Emilie Charton. Analyse longitudinale des données de qualité de vie relative à la santé en cancérologie : vers une standardisation de la méthode du temps jusqu'à détérioration. Médecine humaine et pathologie. Université Bourgogne Franche-Comté, 2020. Français. NNT : 2020UBFCE003 . tel-03212278

HAL Id: tel-03212278
https://theses.hal.science/tel-03212278

Submitted on 29 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ BOURGOGNE FRANCHE-COMTÉ

ÉCOLE DOCTORALE « ENVIRONNEMENTS – SANTÉ »

THÈSE DE DOCTORAT

Présentée et soutenue publiquement

Le 02 Octobre 2020

Par

Emilie CHARTON

Pour obtenir le grade de

Docteur de l'Université Bourgogne Franche-Comté

Discipline : Médecine, Cancérologie, Génétique, Hématologie et Immunologie

Spécialité : Biostatistiques

Analyse longitudinale des données de qualité de vie relative à la santé en cancérologie : vers une standardisation de la méthode du temps jusqu'à détérioration

---

Jury :

Pr Marie-Christine Woronoff-Lemsi, PU-PH, Université Bourgogne Franche-Comté	(directrice de thèse)
Dr Amélie Anota, PhD, Université Bourgogne Franche-Comté	(co-encadrante de thèse)
Pr Bruno Falissard, PU-PH, Université Paris-Sud	(rapporteur)
Pr Francis Guillemin, PU-PH, Université de Lorraine	(rapporteur)
Pr Pascal Hammel, PU-PH, Université Paris VII	(examinateur)
Dr Caroline Bascoul-Mollevi, PhD, Institut du Cancer de Montpellier	(examinateur)
Dr Laurence Collette, PhD, EORTC Headquarters, Bruxelles, Belgique	(examinateur)

A mon père

Remerciements

Mes premières pensées vont naturellement au Pr Franck Bonnetain sans qui cette aventure n'aurait jamais pu débuter. Cette belle aventure qui a commencé devant un poulet frites… Franck, merci de m'avoir mis sur le chemin d'Amélie et d'avoir pu me faire bénéficier de ton savoir à travers elle.

Mes plus grands remerciements sont bien évidemment adressés au Dr Amélie Anota, ma co-encadrante de thèse, qui m'a accompagnée depuis le début et permis de réaliser tous ces travaux. Le 13 avril 2015, date de notre première rencontre, tu m'as permis de découvrir et aimer le domaine de la qualité de vie. Je te remercie de m'avoir tant fait confiance, de m'avoir si bien conseillée et guidée. Tu as toujours su être présente, me supporter mais surtout m'élever pendant toutes ces années. Et nous voilà aujourd'hui à clore un grand chapitre mais aussi à en ouvrir un nouveau.

Je remercie le Pr Marie-Christine Woronoff-Lemsi d'avoir accepté de diriger cette thèse et avoir pris la suite du Pr Franck Bonnetain. J'adresse tous mes remerciements aux Pr Bruno Falissard et Pr Francis Guillemin, mes rapporteurs, de l'honneur qu'ils m'ont fait en acceptant d'examiner et de juger cette thèse. Je tiens à remercier chaleureusement le Pr Pascal Hammel, le Dr Caroline Bascoul-Mollevi et le Dr Laurence Collette pour avoir accepté de faire partie de mon jury.

Je tiens également à remercier mes collègues de l'UMQVC pour le partage de leurs compétences et rigueur scientifique mais également pour leur disponibilité et leurs conseils éclairés. Je suis très heureuse d'avoir croisé votre chemin. Première mention spéciale, pour Morgane, qui m'a tant apportée depuis le début. Tu as toujours su être présente et à l'écoute dans mes moments de doute. Ces quelques mots à ton attention ne suffiront pas pour exprimer tout ce que je ressens. Deuxième mention spéciale, pour les trois JB, qui n'ont jamais cessé de me soutenir et avec qui de vrais liens se sont créés. Enfin, dernière mention spéciale, pour mon binôme de choc, qui m'a toujours aidée, encouragée et fait preuve d'une patience sans limite. Un grand merci pour ta bonne humeur, ta bêtise mais surtout ton humour.

Je tiens aussi à remercier ma mère, ma bibi et Astrid d'avoir eu la gentillesse de faire une relecture de cette thèse. Ma bibi, impossible de ne pas t'écrire quelques mots pour te remercier infiniment pour tout ce que tu m'as apporté depuis toujours. Ta présence, ton amour, ton soutien et tes conseils m'ont toujours été précieux. Je remercie également tous (toutes) mes ami(e)s qui m'ont permis de pouvoir m'évader. Merci à ma belle-famille pour leur immense soutien. Un énorme merci à ma mère, mon frère et ma belle-sœur pour leur amour et qui n'ont jamais cessé de m'épauler et de m'encourager. Moune, je te dois tout.

Enfin, à mon petit loup, pour ton soutien unique qui m'élève et me pousse jour après jour, merci d'être toi, tout simplement.

Table des matières

Abréviations 1

Productions scientifiques 3

Introduction 7

I. Contexte 9

1. La qualité de vie 9

1.1. Histoire 9

1.2. La qualité de vie relative à la santé 10

2. Les instruments de mesure de la qualité de vie relative à la santé 13

2.1. Mode d'administration d'un questionnaire 13

2.2. Les questionnaires 14

2.3. Les questionnaires génériques 14

2.3.1. Le questionnaire SIP 14

2.3.2. Le questionnaire NHP 15

2.3.3. Le questionnaire MOS SF-36 15

2.4. Les questionnaires spécifiques 16

2.4.1. Groupe EORTC 16

2.4.2. Groupe FACT 19

2.4.3. Différences entre les groupes EORTC et FACT 21

2.5. La validation psychométrique des questionnaires de QdV 22

2.5.1. La validité 22

2.5.2. La fiabilité 23

2.5.3. La « *responsiveness* » 24

2.5.4. L'interprétabilité 24

3. Les critères de jugement dans les essais cliniques en cancérologie 25

3.1. La place de la QdV dans les critères de jugement en cancérologie 25

3.2. La place de la QdV au sein des essais cliniques en cancérologie 26

4. Intérêt de la qualité de vie relative à la santé 28

5. Analyse longitudinale des données de QdV 30

5.1. La population d'analyse 30

5.2. La différence minimale cliniquement importante 31

5.3. Méthodes d'analyse longitudinale les plus utilisées 32

5.3.1. Le modèle linéaire à effets mixtes 32

5.3.2. Le temps jusqu'à détérioration d'un score de QdV 34

5.4. Les données manquantes 39

5.4.1. Les différents types de données manquantes 40

5.4.2. Classification des données manquantes 41

5.4.3. Détermination du mécanisme des données manquantes 42

5.4.4. Gestion des données manquantes 43

5.5. L'effet « *response shift* » 46

6. Challenge de l'analyse longitudinale des données de QdV 48

6.1. La différence minimale cliniquement importante 48

6.2. Méthodes d'analyse longitudinale 50

6.3. Les données manquantes 52

6.4. Les essais cliniques en ouvert 54

II. Objectifs 57

III. Résultats 59

1. Revue de la littérature sur la méthode du temps jusqu'à détérioration 59

2. Développement de deux macro-programmes SAS : %TTD et %TUDD 74

3. Investigation de la méthode du temps jusqu'à détérioration 111

3.1. Cohorte de patientes atteintes d'un cancer du sein 111

3.2. Essai de phase II randomisé chez des patients atteints d'un cancer du pancréas 139

Discussion et perspectives 153

Références 161

Annexes 183

Annexe A : Le questionnaire QLQ-C30 183

Annexe B : Le questionnaire FACT-G. 185

Liste des tableaux

Tableau 1. Modules complémentaires spécifiques du groupe EORTC validés. 18

Tableau 2. Dimensions additionnelles spécifiques d'une localisation cancéreuse du groupe FACT. 21

Tableau 3. Différence de conceptualisation des questionnaires de QdV entre les groupes EORTC et FACT d'après Fiteni et al. (2015). 22

Tableau 4. Résumé des différentes définitions du temps jusqu'à détérioration (TJD) et du temps jusqu'à détérioration définitive (TJDD) d'un score de QdV d'après Anota et al. (2013). 36

Liste des figures

Figure 1. Facteurs influençant la QdV d'après Doward et McKenna (2004). 11

Figure 2. Modification du modèle Wilson et Cleary (1995) indiquant le potentiel d'une interaction bidirectionnelle entre les différentes composantes du modèle. 11

Figure 3. Dimension fonctionnement physique relative au questionnaire QLQ-C30. 17

Figure 4. Dimension bien-être physique relative au questionnaire FACT-G. 20

Abréviations

ALD	Lambeau du muscle grand dorsal autologue
AR(1)	First-order autoregressive
ARH(1)	Heterogeneous first-order autoregressive
CSH	Heterogeneous compound symmetry
DATECAN	Definition for the Assessment of Time-to-event Endpoints in CANcer trials
DMCI	Différence minimale cliniquement importante
EORTC	European Organisation for Research and Treatment of Cancer
FACT	Functional Assessment of Cancer Therapy
FDA	Food and Drug Administration
HR	Hazard ratio
IC	Intervalle de confiance
LDI	Lambeau du muscle grand dorsal assisté par implant
LV5FU2	Leucovorine et du fluorouracile
MCAR	Missing completely at random
MAR	Missing at random
MNAR	Missing not at random
MEDLINE	Medical Literature Analysis and Retrieval System Online
MeSH	Medical Subject Headings
MOS SF-36	Medical Outcomes Study Short Form-36
NCI-CTCAE	National Cancer Institute Common Terminology Criteria for Adverse Events
NHP	Nottingham Health Profile
OMS	Organisation Mondiale de la Santé
PROs	Patient-reported outcomes
QdV	Qualité de vie relative à la santé
QFS	Health-related quality of life deterioration-free survival
RMST	Restricted mean survival time
ROC	Receiver operating characteristic
SIP	Sickness Impact Profile
SISAQoL	Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data
TJD	Temps jusqu'à détérioration
TJDD	Temps jusqu'à détérioration définitive
TOI	Trial Outcome Index
UN	Unstructured

Productions scientifiques

Publications

- **Publications principales**

- **Charton E**, Bachet JB, Hammel P, Desramé J, Chibaudel B, Cohen R, Debourdeau P, Dauba J, Lecomte T, Seitz JF, Tournigand C, Aparicio T, Guerin-Meyer V, Taieb J, Volet J, Louvet C, Anota A, Bonnetain F. Impact on health-related quality of life deterioration-free survival of a first-line therapy combining nab-paclitaxel plus either gemcitabine or simplified leucovorin and fluorouracil for patients with metastatic pancreatic cancer: Results of the randomized phase II AFUGEM GERCOR clinical trial. Cancer Medicine 2019;8(11):5079-88.

- **Charton E**, Cuer B, Cottone F, Efficace F, Touraine C, Hamidou Z, Fiteni F, Bonnetain F, Woronoff-Lemsi MC, Bascoul-Mollevi C, Anota A. Time to deterioration in cancer randomized clinical trials for patient-reported outcomes data: a systematic review. Quality of Life Research 2019.

- **Charton E**, Anota A, Bonnetain F, Winters ZE. Time to health-related quality of life score deterioration at 1-year follow-up after types of immediate latissimus dorsi breast reconstructions: a prospective longitudinal cohort study in breast cancer patients. Article soumis.

- **Charton E**, Falcoz A, François E, Touraine C, Bachet JB, Louvet C, Hamidou Z, Bascoul-Mollevi C, Anota A. `%TTD` and `%TUDD`: new SAS macro programs to estimate the time to deterioration for patient-reported outcomes data in oncology. Article en cours de relecture.

- **Autres publications**

- Fenioux C, Louvet C, **Charton E**, Rozet F, Ropert S, Prapotnich D, Barret E, Sanchez-Salas R, Mombet A, Cathala N, Joulia ML, Molitor JL, Henriques J, Bonnetain F, Cathelineau X, Bennamoun M. Switch from abiraterone plus prednisone to abiraterone plus dexamethasone at asymptomatic PSA progression in patients with metastatic castration-resistant prostate cancer. BJU international 2019;123(2):300-6.

- Eberst G, Anota A, Scherpereel A, Mazieres J, Margery J, Greillier L, Audigier-Valette C, Moro-Sibilot D, Molinier O, Lena H, Rivière F, Monnet I, Gounant V, Janicot H, Gervais R, Locher C, **Charton E**, Morin F, Zalcman G, Westeel V. Health-related Quality of Life Impact from Adding Bevacizumab to Cisplatin-Pemetrexed in Malignant Pleural Mesothelioma in the MAPS IFCT-GFPC-0701 Phase III Trial. Clinical Cancer Research 2019:clincanres.2860.018.

- Sparano F, Aaronson NK, Sprangers M, Fayers P, Pusic A, Kieffer JM, Cottone F, Rees J, Pezold M, Anota A, **Charton E**, Vignetti M, Wan C, Blazeby J, Efficace F. Inclusion of older patients with cancer in randomised controlled trials with patient-reported outcomes: a systematic review. BMJ supportive & palliative care 2019.
- Cottone F, Collins GS, Anota A, Sommer K, Giesinger JM, Kieffer JM, Aaronson NK, Van Steen K, **Charton E**, Castagnetti F, Fazi P, Vignetti M, Cella D, Efficace F. Time to health-related quality of life improvement analysis was developed to enhance evaluation of modern anticancer therapies. Journal of Clinical Epidemiology 2020.
- Mouillet G, Efficace F, Thiery-Vuillemin A, **Charton E**, Van Hemelrijck M, Sparano F, Anota A. Investigating the impact of open label design on patient-reported outcome results in prostate cancer randomized controlled trials. Cancer Medicine 2020.
- Cuer B, Bascoul-Mollevi C, Anota A, **Charton E**, Juzyna B, Conroy T, Touraine C. Handling informative dropout in longitudinal analysis of health-related quality of life: Application of three approaches to data from the esophageal cancer clinical trial PRODIGE 5/ACCORD 17. BMC Medical Research Methodology 2020.
- Garnier L, **Charton E**, Falcoz A, Paget-Bailly S, Vernerey D, Jary M, Ducray F, Curtit E. Quality of patient-reported outcomes reporting in randomized controlled trials with glioblastoma patients. Article soumis.
- Anota A, Pozet A, Lemasson H, Cotte FE, Falcoz A, Mouillet G, Eberst G, Guerzider S, **Charton E**, Westeel V. Impact of open-label versus blinded study design on patient-reported outcomes data in randomized clinical trials of immunotherapy in advanced or metastatic cancer patients: a systematic review. Article en cours de relecture.

Communications orales

- **Internationales**

- **Charton E**, Bachet JB, Chibaudel B, Validire P, Hammel P, André T, Louvet C, Anota A, Bonnetain F. Impact on time until deterioration of health-related quality of life of a first-line therapy combining ABI-007 and gemcitabine or simplified LV5FU2 in patients with metastatic pancreatic cancer: results of clinical phase II trial AFUGEM. 23th Annual Conference of the International Society for Quality of Life Research (ISOQOL), 19-22 octobre 2016, Copenhague, Danemark.
- **Charton E**, Anota A, Bonnetain F, Winters ZE. Time to health-related quality of life score deterioration at 1-year follow-up after immediate latissimus dorsi breast reconstructions: a prospective study in breast cancer. 24th Annual Conference of the International Society for Quality of Life Research (ISOQOL), 18-21 octobre 2017, Philadelphie, Pennsylvanie, États-Unis (*finaliste pour le prix étudiant de la meilleure communication orale*).

- **Charton E**, Efficace F, Cottone F, Touraine C, Cuer B, Hamidou Z, Fiteni F, Bonnetain F, Bascoul-Mollevi C, Anota A. Time to health-related quality of life score deterioration in oncology phase III clinical trials: a systematic review. 25th Annual Conference of the International Society for Quality of Life Research (ISOQOL), 24-27 octobre 2018, Dublin, Irlande.

- **Nationales**

- **Charton E**, Bachet JB, Chibaudel B, Validire P, Hammel P, André T, Louvet C, Anota A, Bonnetain F. Results of health-related quality of life of the GERCOR AFUGEM phase II clinical trial in metastatic pancreatic cancer. 10ème Forum du Cancéropôle du Grand-Est, 24-25 novembre 2016, Nancy.

- **Charton E**, Bachet JB, Hammel P, Desramé J, Chibaudel B, Cohen R, Debourdeau P, Dauba J, Lecomte T, Seitz JF, Tournigand C, Aparicio T, Guerin-Meyer V, Taieb J, Volet J, Louvet C, Anota A, Bonnetain F. Impact on quality of life deterioration-free survival in patients with metastatic pancreatic cancer: results of clinical phase II AFUGEM GERCOR trial. Forum des Jeunes Chercheurs, 14-15 juin 2018, Besançon.

- **Charton E**, Cuer B, Touraine C, Hamidou Z, Efficace F, Cottone F, Fiteni F, Bonnetain F, Bascoul-Mollevi C, Anota A. La méthode du temps jusqu'à détérioration d'un score de qualité de vie dans les essais cliniques de phase III en cancérologie : une revue systématique de la littérature. Séminaire Qualité de Vie 5ème édition, 6-7 septembre 2018, Montpellier.

Communications affichées

- **Internationale**

- Anota A, Pozet A, Lefevre C, Lemasson H, Cotte FE, Guerzider S, Mouillet G, Eberst G, **Charton E**, Westeel V. Impact of open-label design on patient-reported outcomes (PROs) data in randomized clinical trials of immuno-oncology (IO) agents in patients with advanced or metastatic cancer: a 10-year systematic literature review (SLR). ESMO Immuno-Oncology Congress, 11-14 décembre 2019, Genève, Suisse.

- **Nationale**

- **Charton E**, Bachet JB, Chibaudel B, Validire P, Hammel P, André T, Louvet C, Anota A, Bonnetain F. Impact sur le temps jusqu'à détérioration de la qualité de vie relative à la santé d'une chimiothérapie de première ligne associant ABI-007 et gemcitabine ou LV5FU2 simplifié chez des patients atteints d'un cancer du pancréas métastatique : résultats de l'essai de phase II AFUGEM. EPICLIN 10/23èmes Journées des Statisticiens des CLCC, 25-27 mai 2016, Strasbourg.

Introduction

Depuis le début des années 1970, le concept de « qualité de vie » a émergé et a suscité un intérêt considérable, tant dans la recherche que dans la pratique clinique (Moons et al., 2006). Gill et Feinstein (1994) ont confirmé cette affirmation en déclarant que « la mesure de la qualité de vie est passée d'une petite entreprise artisanale à une grande entreprise universitaire ».

Le terme « qualité de vie » englobe tous les facteurs autres que la santé (e.g. les données démographiques et l'environnement) qui influencent le bien-être d'un individu. Par la suite, la notion de qualité de vie relative à la santé (QdV) est apparue après la définition de la qualité de vie publiée en 1993 par l'Organisation Mondiale de la Santé (OMS) (WHOQOL, 1993). Cette notion se concentre sur la manière dont la présence d'une maladie et/ou d'un traitement affecte la qualité de vie du patient. La QdV est reconnue comme un concept abstrait et complexe qui intègre trois caractéristiques fondamentales : la multidimensionnalité, la subjectivité et le dynamisme. La multidimensionnalité est une caractéristique essentielle du concept de la QdV. En effet, la QdV inclut au minimum le bien-être physique, psychologique et social, ainsi que les symptômes liés à la maladie et au traitement (Feld, 1995; Leidy et al., 1999). La QdV est une mesure subjective puisque chaque individu a sa propre définition de la QdV (Carr et al., 2001). Cette dernière est appelée un trait latent, puisqu'elle n'est ni observable, ni quantifiable directement. Les mesures de la QdV sont évaluées par l'analyse des réponses aux questionnaires faites par les patients. Enfin, la QdV est également considérée comme dynamique puisqu'un même individu peut évaluer différemment sa QdV au cours du temps (Carr et al., 2001).

La QdV est devenue un critère de jugement de plus en plus important dans les essais cliniques en cancérologie, à côté de critères de jugement tels que la survie globale et la survie sans progression. Elle permet d'évaluer le bénéfice clinique de nouvelles stratégies de traitement pour le patient (Osoba, 2011; Secord et al., 2015). L'intérêt croissant de ce critère peut s'expliquer par la prise de conscience de l'importance de l'impact du traitement sur la QdV des patients, et pas uniquement sur leur corps (Addington-Hall & Kalra, 2001). Cette reconnaissance est nécessaire puisque c'est le patient lui-même qui reçoit le traitement et qui peut ainsi juger et évaluer le bénéfice clinique de par sa perception. De plus, l'*American Society of Clinical Oncology* et la *Food and Drug Administration* (FDA) ont reconnu la QdV comme second critère de jugement principal, en l'absence d'effet sur la survie globale (Beitz et al., 1996; FDA, 2019). Cependant, la QdV est encore bien souvent considérée comme un

critère de jugement secondaire dans les essais.

Dans les essais cliniques en cancérologie, l'objectif de l'évaluation de la QdV est d'étudier l'impact du traitement sur le niveau de QdV du patient au cours du temps. L'analyse longitudinale des données de QdV permettra d'étudier l'évolution de la QdV en tenant compte de la corrélation entre les mesures de QdV d'un même patient. Cependant, l'analyse longitudinale des données de QdV reste complexe (Bonnetain et al., 2016), entre autres parce qu'aucune standardisation des méthodes statistiques n'a encore été proposée. Les diverses manières d'analyser et d'interpréter les données de QdV rendent difficile la comparaison des résultats entre les différents essais cliniques (Bottomley et al., 2016). La méthodologie de l'analyse longitudinale doit être rigoureuse pour éviter tout biais pouvant s'interposer dans les analyses. Des recommandations sont en cours de développement pour pallier le manque de standardisation dans l'analyse et l'interprétation des données de QdV (Bottomley et al., 2016; Coens et al., 2020; Pe et al., 2018).

Ce travail de thèse contribue à la réflexion méthodologique de la standardisation de l'analyse et de l'interprétation des données de QdV, et plus particulièrement sur la méthode du temps jusqu'à détérioration.

Une première partie est consacrée au contexte de la QdV. Elle fait état du concept de la QdV, met en lumière l'intérêt de son évaluation dans les essais cliniques en cancérologie et expose les notions liées à la méthodologie de l'analyse longitudinale en précisant notamment ses différentes problématiques.

De cette partie en découlent les objectifs de mon travail de thèse accompagnés des principaux résultats obtenus au cours de mon doctorat.

Ces derniers font l'objet d'une discussion faisant naître des perspectives de recherches qui viennent conclure ce manuscrit.

I. Contexte

1. La qualité de vie

1.1. Histoire

En 1948, l'OMS définit la santé comme étant « un état de complet bien-être physique, mental et social, et ne consiste pas seulement en une absence de maladie ou d'infirmité » (WHO, 1948). Jusqu'alors, la santé était caractérisée uniquement par l'absence de maladie. Ainsi, avec cette déclaration, l'OMS a élargi la définition de la santé à des aspects plus subjectifs, incluant le bien-être physique, émotionnel, cognitif et social.

Des premiers outils de mesure ont alors été développés pour permettre aux cliniciens d'évaluer les capacités fonctionnelles des patients. L'échelle de Karnofsky (Karnofsky, 1949) est l'un des tout premiers outils de mesure développés. Par la suite, de nouveaux questionnaires ont émergé permettant d'évaluer plusieurs aspects de la qualité de vie, comme par exemple, les fonctionnements physique, émotionnel et social (Walters, 2009). Le *Sickness Impact Profile* (SIP) (Bergner et al., 1976) et le *Nottingham Health Profile* (NHP) (Hunt et al., 1980) font partie de cette nouvelle génération de questionnaires. L'évaluation de plusieurs nouveaux aspects subjectifs de la qualité de vie apparaîtront plus tardivement, avec en particulier le développement du questionnaire *Medical Outcomes Study Short Form-36* (MOS SF-36) en 1992 (Ware & Sherbourne, 1992).

En 1975, « *quality of life* » a été introduit comme un terme *Medical Subject Headings* (MeSH; mot-clé) dans la base de données bibliographiques MEDLINE (*Medical Literature Analysis and Retrieval System Online*) de l'*United States National Library of Medicine*. Ce terme a été accepté comme concept par Index Medicus, index complet d'articles de revues scientifiques médicales, en 1977 (Walters, 2009; Wood-Dauphinee, 1999). Depuis, ce domaine a suscité de l'engouement. Fayers et Jones (1983) ont répertorié plus de 200 articles publiés entre 1978 et 1980 ayant mentionné le terme « *quality of life* » dans le titre. Le nombre de publications dans lesquelles la qualité de vie n'est pas évaluée mais au moins mentionnée ne cesse d'accroître.

En 1993, l'OMS définit la qualité de vie comme « la perception qu'a un individu de sa place dans l'existence, dans le contexte de la culture et du système de valeurs dans lequel il vit, en relation avec ses objectifs, ses attentes, ses normes et ses inquiétudes. Il s'agit donc d'un large champ conceptuel, englobant de manière complexe la santé physique de la personne, son état

psychologique, son niveau d'indépendance, ses relations sociales, ses croyances personnelles et sa relation avec les spécificités de son environnement » (WHOQOL, 1993). Ainsi, la notion de qualité de vie relative à la santé est apparue suite à cette définition. Cette notion se concentre sur les effets de la maladie et en particulier les effets du traitement sur la qualité de vie du patient (Karimi & Brazier, 2016; Leidy et al., 1999).

Malgré toutes ces définitions, une confusion entre les termes santé, qualité de vie et QdV subsiste. Karimi et Brazier (2016) ont démontré que ces trois termes étaient utilisés indifféremment dans la littérature.

Par la suite, le terme « résultats rapportés par le patient » ou « *patient-reported outcomes* » (PROs) est apparu et a été introduit comme un terme MeSH seulement en 2017. Les PROs ont été définis par la FDA comme des mesures de l'état de santé et du traitement perçues par le patient (FDA, 2006). Les PROs évaluent divers domaines tels que : l'état de santé général, les symptômes, l'état fonctionnel, la perception de la santé, la satisfaction du traitement, l'adhérence à un traitement médical ainsi que la QdV.

Dans la suite de ce manuscrit, une différenciation entre les termes qualité de vie et qualité de vie relative à la santé sera faite. Ainsi, l'abréviation QdV fera exclusivement référence à la qualité de vie relative à la santé.

1.2. La qualité de vie relative à la santé

Le concept de la QdV inclut trois caractéristiques fondamentales : multidimensionnalité, subjectivité et dynamisme (Feld, 1995).

Multidimensionnalité. La QdV se structure autour de plusieurs dimensions et incorpore au minimum le bien-être physique, psychologique et social ainsi que les symptômes liés à la maladie et au traitement (Feld, 1995; Leidy et al., 1999). En 2004, Doward et McKenna (2004) ont montré une représentation simplifiée illustrant différents facteurs qui influencent l'évaluation de la QdV d'un individu (*Figure 1*). D'après les auteurs, les facteurs influençant la qualité de vie sont les données démographiques, la culture, l'économie, l'environnement, les relations sociales et la personnalité de l'individu. La maladie et le traitement impactent le niveau symptomatique de l'individu et son état fonctionnel, qui affectent la QdV de l'individu. Enfin, la QdV a une influence sur la qualité de vie. Dans ce modèle, les différents

liens proposés entre les facteurs sont unidirectionnels.

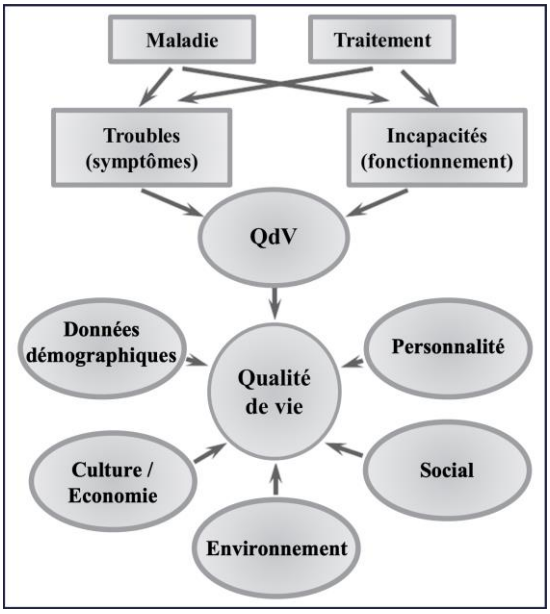


Figure 1. Facteurs influençant la QdV d'après Doward et McKenna (2004).

En 2007, Osoba (2007) propose une version modifiée du modèle conceptuel de Wilson et Cleary (1995). Ce modèle indique cinq niveaux d'influence : les facteurs biologiques et physiologiques (i.e. la fonction des cellules, des organes et des systèmes organiques), le niveau symptomatique (i.e. la perception de l'individu de ses symptômes physiques et psychologiques), l'état fonctionnel (i.e. la capacité de l'individu à effectuer des tâches définies particulières), la perception de l'état de santé général et la QdV globale de l'individu. Tous ces facteurs peuvent également être influencés par des caractéristiques innées et des facteurs environnementaux. Les trois premiers niveaux conduisent à une perception par l'individu de son état de santé général et enfin à une conception de sa QdV globale (*Figure 2*).

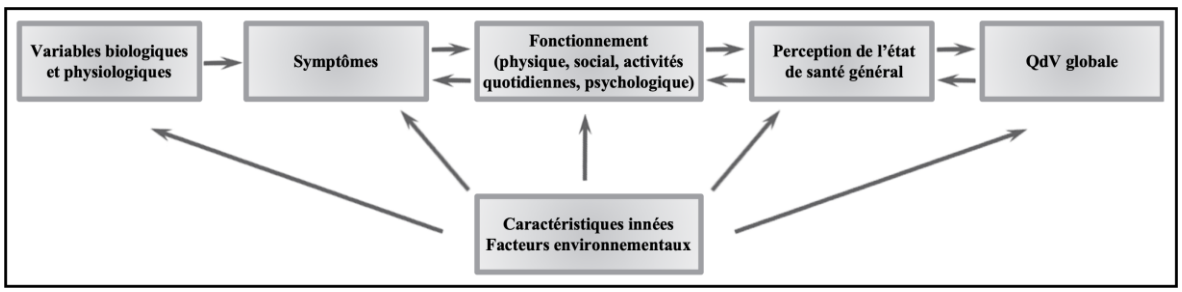


Figure 2. Modification du modèle Wilson et Cleary (1995) indiquant le potentiel d'une interaction bidirectionnelle entre les différentes composantes du modèle.

Initialement, ce modèle considérait seulement des liens unidirectionnels entre les différents niveaux. Ce modèle révisé considère des interactions bidirectionnelles entre le niveau symptomatique, l'état fonctionnel, la perception de l'état de santé général et la QdV globale. Ainsi, la QdV globale peut avoir un impact sur la perception de l'état de santé général qui, à son tour, peut affecter l'état fonctionnel et enfin, avoir un effet sur l'expression de certains symptômes.

Subjectivité. La QdV n'est ni observable, ni quantifiable directement (appelée un trait latent), d'où l'aspect subjectif de la mesure. La QdV est exclusivement évaluée du point de vue de l'individu (Osoba, 1994). Une variation entre les individus (inter-individuelle), au regard de la définition de la QdV existe (Carr et al., 2001). En effet, chaque individu a sa propre définition de la QdV, influencée par ses motivations personnelles, ses émotions, son ressenti, sa culture et son environnement (Carr & Higginson, 2001). Par exemple, deux patients ayant le même état de santé n'auront pas forcément le même ressenti face à la douleur, et donc, n'auront pas la même perception de leur QdV. Tous ces éléments viennent corroborer la définition du terme « subjectif » issue du dictionnaire Larousse : « se dit de ce qui est individuel et susceptible de varier en fonction de la personnalité de chacun ».

Dynamisme. La subjectivité de la mesure entraîne l'aspect dynamique de la QdV. En effet, il est également possible d'observer une variation au sein d'un même individu (intra-individuelle), au regard de la définition de la QdV (Carr et al., 2001). Un individu peut évaluer différemment sa QdV au cours du temps, en fonction de ses attentes et de ses espérances de santé qui peuvent évoluer, mais aussi en raison de facteurs influençant la QdV comme décrits plus haut. Par exemple, avant le diagnostic d'une maladie, un patient peut accorder plus d'importance à la composante physique. Après diagnostic, ce même patient pourrait revoir ses priorités et se tourner davantage vers une dimension spirituelle ou familiale.

2. Les instruments de mesure de la qualité de vie relative à la santé

Les mesures de la QdV reposent sur les jugements que les patients portent sur leur propre état de santé (Fiteni et al., 2015). Elles sont évaluées par l'analyse des réponses aux questionnaires faites par les patients. Il est préférable que l'évaluation soit effectuée par le patient lui-même afin de prendre en compte la subjectivité de la mesure (Addington-Hall & Kalra, 2001). Cependant, dans certains cas, le patient n'est pas dans la capacité de remplir lui-même les questionnaires. Différents modes d'administration ont donc été mis en place.

2.1. Mode d'administration d'un questionnaire

Le choix du mode d'administration d'un questionnaire dépend de différents facteurs. On distingue trois modes d'administration : l'auto-évaluation, l'hétéro-évaluation et le proxy.

L'auto-évaluation correspond à une évaluation effectuée par le patient lui-même. Ce mode d'administration suppose que le patient est apte à fournir des informations valables et qu'il est en mesure d'évaluer sa propre QdV (Addington-Hall & Kalra, 2001). Cependant, dans certaines conditions, une auto-évaluation peut être difficile voire impossible. Par exemple, en raison d'une incapacité à répondre, d'une incapacité à fournir des informations valables, de troubles cognitifs, de symptômes ou de la sévérité de leur maladie, de nombreux patients sont incapables d'évaluer leur QdV. Ainsi lorsque l'auto-évaluation semble difficile, une hétéro-évaluation peut être effectuée. L'hétéro-évaluation consiste à énoncer le questionnaire au patient par une tierce personne, qui le remplira à la place du patient. Enfin, lorsque l'évaluation de la QdV est impossible par le patient, des questionnaires ont été développés pour être renseignés par un aidant, faisant partie de la sphère familiale, amicale ou médicale. Ce mode d'administration est appelé une mesure proxy, ce qui correspond à une évaluation par procuration (Pickard & Knight, 2005). La mesure proxy introduit une part du point de vue de l'aidant. Ainsi, l'aidant pourrait sous-estimer ou surestimer certains aspects de la QdV du patient (Conde-Sala et al., 2009). Cependant, ce mode d'administration peut apporter des informations complémentaires sur le patient en obtenant le point de vue de l'aidant.

2.2. Les questionnaires

Les questionnaires de QdV sont généralement composés de plusieurs questions, appelées aussi items. Les modalités de réponse aux items peuvent être binaires (réponses du type « oui » ou « non), qualitatives ordinales aussi appelées échelle de Likert (réponses du type « pas du tout », « un peu », « assez » ou « beaucoup ») ou représentées par des échelles visuelles analogiques (échelles allant de 0 à 10 ou parfois de 0 à 100). Les items sont regroupés en une ou plusieurs dimensions représentant différents aspects de la QdV (Mercier & Schraub, 2005). Ces dimensions peuvent être évaluées suivant un ou plusieurs items. Un score est généralement calculé pour chaque dimension, reflétant le niveau du patient pour la dimension correspondante.

Les questionnaires peuvent être génériques ou spécifiques d'une maladie (Mercier & Schraub, 2005). Les questionnaires génériques sont administrables auprès de sujets malades ou non malades, c'est-à-dire quel que soit leur état de santé. Ces questionnaires permettent de comparer la QdV de sujets présentant des pathologies très différentes. Les questionnaires spécifiques ont été développés pour évaluer la QdV pour une population précise, ayant une pathologie bien définie (cardiologie, neurologie, cancérologie, etc.) ou un symptôme particulier (douleur, fatigue, insomnie, etc.).

2.3. Les questionnaires génériques

2.3.1. Le questionnaire SIP

Le questionnaire SIP est un questionnaire générique développé en 1976 (Bergner et al., 1976) évaluant l'effet de la maladie sur le fonctionnement physique et émotionnel. Ce questionnaire contient 136 items (questions) ayant des modalités de réponse binaires (i.e. « oui » ou « non »). Il a été validé en langue française (Chwalow et al., 1992) et évalue 12 dimensions : le sommeil et le repos, l'humeur/l'état psychique, les soins corporels, les occupations à la maison, les déplacements à l'extérieur, la relation avec les proches, la marche, la réflexion/la mémoire et la concentration, l'aptitude à communiquer, le travail, les loisirs et l'alimentation. Les dimensions « marche », « déplacements à l'extérieur » et « soins corporels » sont synthétisables en un score résumé physique et les 9 autres en un score résumé psychosocial. Un score global peut également être généré à partir de tous les items. Chaque score est multiplié par 100 afin d'obtenir un pourcentage, où 0 correspond à des individus en parfaite

santé et 100 à des individus avec un dysfonctionnement maximal.

2.3.2. Le questionnaire NHP

Le questionnaire NHP est un questionnaire générique développé en 1980 (Hunt et al., 1980) quantifiant la perception de l'état de santé sur le plan physique, émotionnel et social. La version finale de ce questionnaire se décompose en deux parties indépendantes. La première partie contient 38 items et a été validée en langue française (Bucquet et al., 1990). La deuxième partie contient 7 items. Les modalités de réponse aux items du questionnaire sont binaires (i.e. « oui » ou « non »). La première partie du questionnaire évalue 6 dimensions : la mobilité, l'isolation sociale, les réactions émotionnelles, la douleur, l'énergie et le sommeil. Au sein de chaque dimension, chaque item est pondéré selon la sévérité perçue. Ainsi, par dimension, le score est calculé en faisant la somme des produits des réponses aux items (« oui » représentant « 1 » et « non » représentant « 0 ») multipliée par le coefficient correspondant. Les scores varient de 0 (absence de handicap) à 100 (handicap majeur), correspondant au degré de difficulté que perçoit l'individu dans la dimension correspondante. La deuxième partie du questionnaire explore les problèmes quotidiens rencontrés (travail rémunéré, travaux domestiques, vie sociale, relations personnelles, vie sexuelle, centres d'intérêt/hobbies et vacances). Cette partie est facultative, elle peut donc être omise sans affecter la validité et la fiabilité globales de l'instrument.

2.3.3. Le questionnaire MOS SF-36

Le questionnaire MOS SF-36 est un questionnaire générique développé en 1992 (Ware & Sherbourne, 1992) évaluant l'état de santé. Ce questionnaire contient 36 items, construits sur des échelles variant entre deux et six modalités de réponse (e.g. « oui » ou « non » et « pas du tout », « un petit peu », « moyennement », « beaucoup » ou « énormément »). Le questionnaire MOS SF-36 a été traduit et adapté en langue française (Leplège et al., 1998). Il évalue 9 dimensions de QdV : l'activité physique, les limitations dues à l'état physique, les douleurs physiques, la santé perçue, la vitalité, la vie et les relations avec les autres, la santé psychique, les limitations dues à l'état psychique et l'évolution de la santé perçue. Les procédures de calculs des scores de ce questionnaire sont reportées dans le manuel de Ware et al. (1993). Les scores varient de 0 (pire QdV) à 100 (meilleure QdV). Lorsque des données manquantes sont constatées, le manuel de scoring recommande qu'un score soit calculé si

l'individu a répondu à au moins la moitié des items, dans une dimension à plusieurs items. Dans le cas d'une dimension possédant un nombre impair d'items, l'individu doit avoir répondu à plus de la moitié des items. Les quatre premières dimensions sont synthétisables en un score résumé physique et les quatre suivantes en un score résumé psychique. Ainsi l'évolution de la santé perçue ne rentre pas dans les scores résumés.

2.4. Les questionnaires spécifiques

Deux grands groupes se distinguent dans le développement d'auto-questionnaires spécifiques couramment utilisés en cancérologie : le groupe QdV de l'*European Organisation for Research and Treatment of Cancer* (EORTC) et le groupe américain *Functional Assessment of Cancer Therapy* (FACT).

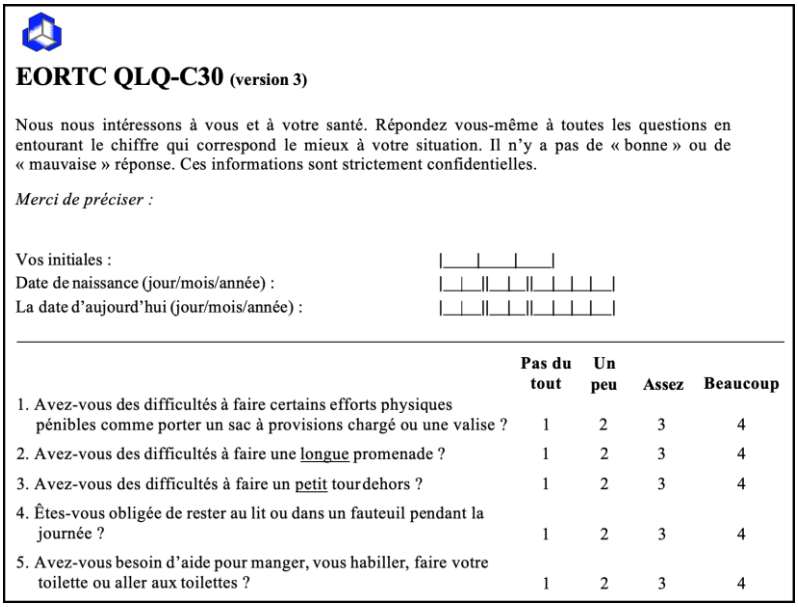
2.4.1. Groupe EORTC

Le groupe EORTC a commencé par développer un questionnaire spécifique du cancer, nommé QLQ-C30 (Aaronson et al., 1993), qui a été validé dans plus de 110 langues différentes dont le français. Ce questionnaire contient 30 items dont 28 items construits sur une échelle à quatre modalités de réponse (i.e. « pas du tout », « un peu », « assez » ou « beaucoup »). Les 2 derniers items sont construits sur une échelle à sept niveaux (allant de 1 « très mauvais » à 7 « excellent ») (*Annexe A*). Ce questionnaire évalue 15 dimensions de QdV qui sont les suivantes :

- 1 dimension de QdV/santé globale :
    - QdV/santé globale (items 29 et 30)
- 5 dimensions fonctionnelles :
    - Fonctionnement physique (items 1 à 5)
    - Fonctionnement de rôle (activités quotidiennes, items 6 et 7)
    - Fonctionnement émotionnel (items 21 à 24)
    - Fonctionnement cognitif (items 20 et 25)
    - Fonctionnement social (items 26 et 27)
- 9 dimensions symptomatiques :
    - Fatigue (items 10, 12 et 18)
    - Nausée et vomissement (items 14 et 15)

- Douleur (items 9 et 19)
- Dyspnée (item 8)
- Insomnie (item 11)
- Perte d'appétit (item 13)
- Constipation (item 16)
- Diarrhée (item 17)
- Difficultés financières (item 28)

Pour chaque dimension, un score est calculé. Par exemple, les cinq premiers items de ce questionnaire permettent de calculer un score de fonctionnement physique (*Figure 3*). Le groupe EORTC a créé un manuel de scoring contenant les procédures des calculs des différents scores (Fayers et al., 2001). Les scores varient de 0 (pire) à 100 (meilleur) pour les dimensions QdV/santé globale et fonctionnelles, et de 0 (meilleur) à 100 (pire) pour les dimensions symptomatiques. Le groupe EORTC recommande un pourcentage maximum de 50% d'items manquants par dimension pour pouvoir déterminer le score. Au-delà de ce pourcentage, le score ne pourra pas être calculé. Les items manquants ne sont pas pris en compte dans le calcul du score, c'est-à-dire que l'on procède à une imputation simple par la moyenne, considérant que l'item manquant ne diffère pas en moyenne des autres items. De plus, une version abrégée de 15 éléments du QLQ-C30 a été développée et validée pour les soins palliatifs, le questionnaire QLQ-C15-PAL (Groenvold et al., 2006).



EORTC QLQ-C30 (version 3)

Nous nous intéressons à vous et à votre santé. Répondez vous-même à toutes les questions en entourant le chiffre qui correspond le mieux à votre situation. Il n'y a pas de « bonne » ou de « mauvaise » réponse. Ces informations sont strictement confidentielles.

Merci de préciser :

Vos initiales : |___|___|___|

Date de naissance (jour/mois/année) : |__|__||__|__||_|_|_|_|

La date d'aujourd'hui (jour/mois/année) : |__|__||__|__||_|_|_|_|

	Pas du tout	Un peu	Assez	Beaucoup
1. Avez-vous des difficultés à faire certains efforts physiques pénibles comme porter un sac à provisions chargé ou une valise ?	1	2	3	4
2. Avez-vous des difficultés à faire une longue promenade ?	1	2	3	4
3. Avez-vous des difficultés à faire un petit tour dehors ?	1	2	3	4
4. Êtes-vous obligée de rester au lit ou dans un fauteuil pendant la journée ?	1	2	3	4
5. Avez-vous besoin d'aide pour manger, vous habiller, faire votre toilette ou aller aux toilettes ?	1	2	3	4

Figure 3. Dimension fonctionnement physique relative au questionnaire QLQ-C30.

Par ailleurs, le groupe EORTC a également développé des modules complémentaires spécifiques, à administrer conjointement avec le questionnaire QLQ-C30 (questionnaires additionnels), ainsi que des questionnaires pouvant être utilisés indépendamment du questionnaire QLQ-C30. Ces questionnaires peuvent être spécifiques d'une localisation cancéreuse, d'un symptôme particulier ou d'un domaine particulier. Le calcul des scores des différents questionnaires se fait selon le même manuel de scoring que pour le questionnaire QLQ-C30. Le *Tableau 1* ci-dessous présente les différents modules complémentaires spécifiques validés existants.

Tableau 1. Modules complémentaires spécifiques du groupe EORTC validés.

Nom	Spécificité	Nombre d'items	Nombre de dimensions	Référence (auteur, année)
QLQ-BIL21	Cancer des voies biliaires	21	5	Kaupp-Roberts et al. (2016)
QLQ-BM22*	Métastases osseuses	22	4	Chow et al. (2012)
QLQ-BN20	Cerveau	20	11	Taphoorn et al. (2010)
QLQ-BR23	Cancer du sein	23	8	Sprangers et al. (1996)
QLQ-BRECON23[#]	Reconstruction mammaire	23	9	Winters et al. (2018)
QLQ-CR29	Cancer colorectal	35	23	Whistance et al. (2009)
QLQ-CX24	Cancer du col de l'utérus	24	9	Greimel et al. (2006)
QLQ-ELD14	Patients âgés atteints d'un cancer	14	7	Wheelwright et al. (2013)
QLQ-EN24	Cancer de l'endomètre	24	13	Greimel et al. (2011)
QLQ-FA12	Fatigue liée au cancer	12	5	Weis et al. (2017)
QLQ-GINET21	Tumeurs neuroendocrines gastro-intestinales	21	10	Yadegarfar et al. (2013)
QLQ-HCC18	Carcinome hépatocellulaire	18	8	Chie et al. (2012)
QLQ-H&N35	Cancer tête et cou (ancienne version)	35	18	Bjordal et al. (1999)
QLQ-H&N43	Cancer tête et cou (nouvelle version)	43	19	Singer et al. (2015)
QLQ-INFO25	Information	25	12	Arraras et al. (2010)
QLQ-LC13	Cancer du poumon	13	11	Bergman et al. (1994)
QLQ-LMC21	Cancer colorectal avec métastases au foie	21	13	Blazeby et al. (2009)
QLQ-MY20	Myélome multiple	20	4	Cocks et al. (2007)
QLQ-OES18	Cancer de l'œsophage	18	10	Blazeby et al. (2003)
QLQ-OG25	Cancer œsophago-gastrique	25	16	Lagergren et al. (2007)
QLQ-OH15	Santé bucco-dentaire	15	6	Hjermstad et al. (2016)
QLQ-OV28	Cancer de l'ovaire	28	8	Greimel et al. (2003)
QLQ-PR25	Cancer de la prostate	25	6	van Andel et al. (2008)
QLQ-PRT20	Rectite radique	20	5	Halkett et al. (2018)
QLQ-STO22	Cancer gastrique	22	9	Blazeby et al. (2004)
QLQ-TC26	Cancer des testicules	26	12	Sztankay et al. (2018)

Tous les modules sont utilisés conjointement avec le questionnaire QLQ-C30 sauf indication contraire; *Module administrable conjointement avec les questionnaires QLQ-C30 ou QLQ-C15-PAL; [#]Module administrable conjointement avec les questionnaires QLQ-C30 et QLQ-BR23.

Le groupe EORTC ne cesse de développer de nouveaux questionnaires. Par exemple, le module QLQ-ANL27 pour le cancer anal (Sodergren et al., 2018) ou encore le module QLQ-MEL38 spécifique des patients atteints d'un mélanome (Winstanley et al., 2015). Ces deux questionnaires sont en cours de validation. Les questionnaires déjà validés sont régulièrement mis à jour, que cela soit au niveau du nombre de questions (i.e. suppression et/ou ajout de questions), ou au niveau de la formulation. Par exemple, le module spécifique QLQ-H&N35

pour des patients atteints d'un cancer tête et cou, avait initialement été développé en 1999 (Bjordal et al., 1999) et a été révisé et remplacé par le module QLQ-H&N43 en 2015 (Singer et al., 2015). Ce questionnaire a été mis à jour au regard du nombre de questions, passant de 35 à 43 items et également au regard de la formulation de ces questions (e.g. la question « Avez-vous toussé ? » est devenue « Avez-vous eu des problèmes de toux ? »). Dans cette nouvelle version, certaines questions ont également été supprimées pour être remplacées par d'autres plus adaptées à la prise en charge actuelle. Ainsi, les dimensions ont été revues en conséquence, en modifiant les dimensions existantes, en combinant ensemble d'anciennes dimensions ou en créant de nouvelles.

Un guide pour le développement des modules au sein du groupe EORTC a été proposé et se décompose en quatre phases (EORTC Quality of Life Group, April 2011). La phase 1 permet de générer les domaines de QdV pertinents par le biais d'une revue de la littérature, d'une interview auprès des professionnels de santé et des patients. La phase 2 consiste à convertir les domaines de QdV pertinents en question de telle sorte qu'ils soient compatibles avec le QLQ-C30 et qu'ils soient uniformes avec les modules déjà existants. La rédaction se fait désormais avec l'aide de la banque d'items développée par le groupe EORTC (https://www.eortc.be/itemlibrary). La phase 3 consiste à faire un test pilote du questionnaire provisoire afin d'identifier et de résoudre les problèmes d'administration (formulation et ordre des questions) et de déterminer si des questions sont manquantes ou redondantes. Ainsi, le nouveau module conjointement au questionnaire QLQ-C30 est administré aux patients correspondant à la population cible du module en cours de développement. Par la suite, la population cible répond à un entretien semi-structuré visant à demander si des questions étaient non pertinentes et si des questions supplémentaires pertinentes sont manquantes dans le questionnaire provisoire. Sur la base de cette phase de pré-test, le questionnaire provisoire peut nécessiter une adaptation. Enfin, la phase 4 permet de valider les propriétés psychométriques en testant le module auprès d'un grand groupe international de patients afin de déterminer son acceptabilité, sa fiabilité, sa validité, sa réactivité et son applicabilité interculturelle.

2.4.2. Groupe FACT

Le groupe FACT a également développé un questionnaire spécifique du cancer, nommé le FACT-General (FACT-G) (Cella et al., 1993) qui a été validé dans de nombreuses langues

dont le français (Conroy et al., 2004). Ce questionnaire contient 27 items construits sur une échelle à cinq modalités de réponse (i.e. « pas du tout », « un peu », « moyennement », « beaucoup » ou « énormément ») (*Annexe B*). Il évalue 4 dimensions de bien-être : le bien-être physique, social/familial, émotionnel et fonctionnel. Par exemple, les sept premiers items de ce questionnaire permettent de calculer le score de bien-être physique (*Figure 4*). Les scores varient de 0 à 28 pour toutes les dimensions, hormis la dimension bien-être émotionnel qui varie de 0 à 24. Un score élevé représentera un niveau élevé de bien-être. Lorsque des données manquantes sont constatées, le groupe FACT recommande un pourcentage maximum de 50% d'items manquants par dimension afin de pouvoir calculer le score correspondant. Au-delà de ce pourcentage, le score ne pourra pas être calculé. De plus, un score FACT-G global est calculé comme la somme des dimensions de bien-être. Le score global peut être calculé si le pourcentage maximum de données manquantes est inférieur à 80% et varie de 0 à 108.

	BIEN-ÊTRE PHYSIQUE	**Pas du tout**	**Un peu**	**Moyen-nement**	**Beau-coup**	**Énormé-ment**
GP1	Je manque d'énergie	0	1	2	3	4
GP2	J'ai des nausées	0	1	2	3	4
GP3	À cause de mon état physique, j'ai du mal à répondre aux besoins de ma famille	0	1	2	3	4
GP4	J'ai des douleurs	0	1	2	3	4
GP5	Je suis incommodé(e) par les effets secondaires du traitement	0	1	2	3	4
GP6	Je me sens malade	0	1	2	3	4
GP7	Je suis obligé(e) de passer du temps allongé(e)	0	1	2	3	4

Figure 4. Dimension bien-être physique relative au questionnaire FACT-G.

Par ailleurs, le groupe FACT a développé des dimensions additionnelles à ajouter au sein du questionnaire FACT-G. Ces dimensions peuvent être spécifiques à la localisation cancéreuse, au traitement et à un symptôme particulier. Par exemple, le questionnaire FACT-B spécifique du cancer du sein développé par Brady et al. (1997) contient 27 items mesurant les dimensions relatives au questionnaire FACT-G, ainsi que 10 items additionnels évaluant une dimension spécifique au cancer du sein. Un score FACT-B global est calculé comme la somme des dimensions de bien-être du FACT-G et de la dimension additionnelle spécifique au cancer du sein. Le groupe FACT a proposé des indices nommés « *Trial Outcome Index* »

(TOI) selon les questionnaires correspondant à un score résumé. En effet, le FACT-B TOI correspond à un score résumé des dimensions bien-être physique, fonctionnel et de la dimension spécifique au cancer du sein. Le groupe FACT a également développé des questionnaires correspondant à un sous-ensemble d'items d'un autre questionnaire. Par exemple, le questionnaire FBSI (*FACT-Breast Symptom Index*) est une version abrégée de 8 éléments du questionnaire FACT-B (Yost et al., 2005). Le *Tableau 2* ci-dessous présente différents exemples de dimensions additionnelles au FACT-G spécifiques d'une localisation cancéreuse.

Tableau 2. Dimensions additionnelles spécifiques d'une localisation cancéreuse du groupe FACT.

Nom	Localisation cancéreuse	Nombre d'items	Référence (auteur, année)
FACT-B	Sein	37	Brady et al. (1997)
FACT-Bl	Vessie	40	Degboe et al. (2019)
FACT-Br	Cerveau	50	Weitzner et al. (1995)
FACT-C	Colorectal	37	Ward et al. (1999)
FACT-E	Œsophage	44	Darling et al. (2006)
FACT-Hep	Hépato-biliaire	45	Heffernan et al. (2002)
FACT-L	Poumon	37	Cella et al. (1995)
FACT-Leu	Leucémie	44	Cella et al. (2012)
FACT-M	Mélanome	43	Cormier et al. (2008)
FACT-NP	Nasopharynx	43	Tong et al. (2009)
FACT-O	Ovaire	39	Basen-Engquist et al. (2001)
FACT-P	Prostate	39	Esper et al. (1997)
FACT-V	Vulve	46	Janda et al. (2005)

Tous les questionnaires évaluent 5 dimensions (quatre dimensions de bien-être du FACT-G et une dimension spécifique de la localisation cancéreuse) et 3 dimensions globales (un score FACT-G global, un score TOI et un score global spécifique à la localisation cancéreuse).

2.4.3. Différences entre les groupes EORTC et FACT

Les groupes EORTC et FACT n'ont pas la même conceptualisation des questionnaires de QdV. Par exemple, le groupe EORTC développe par localisation cancéreuse des modules complémentaires qui sont administrables conjointement au questionnaire QLQ-C30. Le groupe FACT, quant à lui, ajoute simplement une dimension additionnelle au questionnaire FACT-G. Au sein des questionnaires du groupe FACT, les items sont séparés par domaine de QdV. Le patient peut donc directement visualiser les domaines qui seront mesurés par le biais du questionnaire contrairement au groupe EORTC où les domaines ne sont pas énoncés. Les différences de conceptualisation des questionnaires de QdV entre ces deux groupes sont résumées dans le *Tableau 3* ci-dessous.

Tableau 3. Différence de conceptualisation des questionnaires de QdV entre les groupes EORTC et FACT d'après Fiteni et al. (2015).

	EORTC	**FACT**
Domaines	Centrés sur les symptômes et dimensions fonctionnelles	Dimension de bien-être et dimensions de préoccupations additionnelles
Structure	Items non séparés par domaine Domaines de QdV évalués non énoncés	Items séparés par domaine Items précédés par une étiquette indiquant la dimension mesurée ex : « bien-être physique »
Échelles	Uni- et multi-items	Multi-items
Modalité de réponse par item	4 modalités	5 modalités
Questions formulées	À la seconde personne ex : « Avez-vous eu des nausées ? »	À la première personne ex : « J'ai des nausées »
Par dimension	Items formulés dans le même sens	Nécessité d'inverser certains items pour le calcul du score
Score de QdV globale	Évalué par deux items du QLQ-C30	Moyenne des scores obtenus pour chaque dimension
Par localisation cancéreuse	1 module complémentaire Module à administrer conjointement avec le QLQ-C30 Plusieurs dimensions spécifiques de la localisation	1 dimension additionnelle au questionnaire FACT-G 1 seul questionnaire à administrer 1 seule dimension spécifique de la localisation
Développement	Simultanément dans plusieurs langues et cultures	Dans une langue, puis de façon séquentielle adaptation à d'autres langues et cultures

2.5. La validation psychométrique des questionnaires de QdV

Le développement d'un nouvel instrument de QdV doit satisfaire des propriétés psychométriques, dans le but de mesurer correctement et de façon robuste la QdV (Fayers & Machin, 2000; Walters, 2009). Ces propriétés sont principalement la validité, la fiabilité, la « *responsiveness* » et l'interprétabilité. Ces propriétés doivent également être respectées lors d'une adaptation transculturelle ou lors de la réduction d'un questionnaire déjà existant.

2.5.1. La validité

La validité est la capacité avec laquelle un instrument mesure ce qu'il est censé évaluer. Il existe plusieurs types de validité.

La validité du contenu (« *content validity* ») permet de vérifier si les items sont pertinents et couvrent de manière exhaustive le domaine d'intérêt. Cette validité concerne également l'adéquation du contenu de l'instrument, en termes de nombre et de scope des items. Ce concept consiste à un examen critique de la structure de l'instrument.

La validité apparente (« *face validity* »), souvent considérée comme une forme de validité de contenu, évalue si les items de l'instrument couvrent clairement le domaine d'intérêt, sans ambiguïté et avec crédibilité. Ce concept concerne la revue critique d'un instrument après

avoir été construit. Cette dernière est effectuée par le biais de l'opinion d'experts médicaux et de patients.

La validité du critère (« *criterion validity* ») implique l'évaluation de la concordance du nouvel instrument par rapport à un critère externe, comme par exemple un instrument déjà validé, évaluant le même concept ou les mêmes dimensions de QdV. La validité du critère est généralement divisée en deux types : la validité simultanée (« *concurrent validity* ») et la validité prédictive (« *predictive validity* »). La validité simultanée concerne la corrélation entre le nouvel instrument et le critère externe, tous deux mesurés en même temps. Lorsque le nouvel instrument est corrélé avec le critère externe, c'est à ce moment qu'intervient la validité prédictive. Elle concerne la capacité de l'instrument à prédire l'état de santé futur, par le biais des scores du nouvel instrument.

La validité de structure (« *construct validity* ») aborde la question de savoir si l'instrument a bien mesuré ce qui devait l'être. Elle est généralement divisée en trois types : la validité de groupes connus (« *known-groups validity* »), la validité convergente (« *convergent validity* ») et la validité divergente (« *divergent validity* »). La validité de groupes connus est basée sur l'hypothèse que certains groupes de sujets pourraient avoir un score différent par rapport à d'autres groupes. Ainsi, le nouvel instrument devrait être sensible à ces différences. La validité convergente, quant à elle, consiste à montrer que la dimension de QdV établie est corrélée de façon satisfaisante avec toutes les autres dimensions qui, selon la théorie, devraient y être liées. Enfin, la validité divergente reconnaît que certaines dimensions de QdV devraient être relativement indépendantes et que leurs corrélations devraient être faibles.

2.5.2. La fiabilité

La fiabilité consiste à déterminer si l'instrument de mesure permet de donner des résultats cohérents et reproductibles. La cohérence interne (« *internal consistency reliability* ») permet de s'assurer que les items qui constituent une dimension soient homogènes entre eux. La reproductibilité, quant à elle, correspond à la stabilité des scores de QdV au cours du temps. Une population qui n'a connu aucun changement de santé ou de QdV devrait avoir des scores de QdV inchangés. Cette reproductibilité est mesurée selon la méthode « *test-retest reliability* » et consiste à administrer à un patient plusieurs fois le même instrument. Ainsi, les réponses doivent être en cohérence pour ce patient.

2.5.3. La « *responsiveness* »

La « *responsiveness* » est la mesure dans laquelle un instrument est capable de détecter un changement cliniquement important du statut de QdV. En effet, un instrument doit non seulement être fiable, produisant des résultats reproductibles lorsque l'état de santé ou de QdV d'un patient est stable, mais il doit aussi répondre aux changements de l'état du patient. Par exemple, si la progression de la maladie entraîne une détérioration de la QdV d'un patient, l'instrument de mesure doit être suffisamment sensible pour détecter ce changement dans l'état de santé du patient. De même, si deux groupes de patients diffèrent dans leur QdV, l'instrument doit être suffisamment sensible pour détecter ce changement.

2.5.4. L'interprétabilité

L'interprétabilité des scores obtenus sur un instrument est un autre critère à considérer. Ce critère n'est pas considéré comme une propriété de mesure, mais comme une caractéristique importante (Mokkink et al., 2010). Les instruments peuvent être capables de détecter des changements, mais il est important de déterminer si ces changements sont significatifs. Ainsi, une des approches pour interpréter ces résultats consiste à déterminer la plus petite différence ou variation des scores de QdV qui serait considérée comme ayant un sens clinique pour le patient, appelée la différence minimale cliniquement importante (Jaeschke et al., 1989). La notion de différence minimale cliniquement importante sera détaillée au paragraphe 5.2.

3. Les critères de jugement dans les essais cliniques en cancérologie

Les critères de jugement sont des mesures cliniques et biologiques pour l'évaluation et le développement des stratégies thérapeutiques. En cancérologie, les critères de jugement peuvent être regroupés en deux catégories : les critères « centrés sur le patient » permettant de s'assurer d'un bénéfice clinique pour le patient et ceux « centrés sur la tumeur » permettant d'observer l'évolution de la maladie (Bonnetain, 2010; Fiteni et al., 2014). Les critères de jugement « centrés sur le patient » incluent par exemple la survie globale (i.e. le délai entre la date de randomisation et le décès) et la QdV. Ceux « centrés sur la tumeur » incluent entre autres la survie sans progression (i.e. le délai entre la date de randomisation et la première progression tumorale ou le décès).

En cancérologie, la survie globale demeure le critère de jugement de référence dit « *gold standard* » démontrant un bénéfice clinique pour le patient. Cependant, les traitements devenant de plus en plus efficaces pour de nombreux types de cancers entraînent la nécessité d'inclure un nombre plus important de patients et la nécessité de suivre les patients à long terme afin d'observer un nombre suffisant de décès, permettant d'atteindre la puissance statistique prévue (Fiteni et al., 2014). Ainsi, une alternative serait d'utiliser des critères de jugement dits intermédiaires centrés sur la tumeur qui permettraient une évaluation plus précoce et qui seraient utilisés comme critères substitutifs de la survie globale. Ces critères « centrés sur la tumeur » ont rarement été validés comme critères substitutifs de la survie globale. Ils reposent sur de multiples définitions d'événements non standardisées (Bellera et al., 2013). Cependant, le projet DATECAN (*Definition for the Assessment of Time-to-event Endpoints in CANcer trials*) a permis de fournir des recommandations visant à standardiser les définitions couramment utilisées pour les critères « centrés sur la tumeur » dans les principales localisations cancéreuses (Bellera et al., 2015; Bonnetain et al., 2014; Cohen et al., 2020; Gourgou-Bourgade et al., 2015; Kramar et al., 2015). Néanmoins, ces critères ne reflètent pas directement le possible bénéfice clinique pour le patient. Dans ce contexte, la QdV constitue un critère de jugement d'intérêt pour évaluer le bénéfice clinique perçu par le patient.

3.1. La place de la QdV dans les critères de jugement en cancérologie

La QdV est reconnue comme second critère de jugement principal par l'*American Society of*

Clinical Oncology et la FDA en l'absence d'effet sur la survie globale (Beitz et al., 1996; FDA, 2019). Cependant, la QdV est souvent considérée comme un critère de jugement secondaire.

Dans certaines situations, elle semblerait pertinente comme critère de jugement principal ou comme co-critère de jugement principal, par exemple, en oncogériatrie (Wildiers et al., 2012) ou en situation palliative (Pazart et al., 2014). Chez les patients âgés atteints d'un cancer, la QdV pourrait être plus importante que la durée de vie (Shrestha et al., 2019), compte tenu de leurs potentielles comorbidités, fragilité et de leur espérance de vie naturellement plus limitée. En fin de vie, elle est considérée comme un critère de jugement majeur. L'objectif n'est pas de démontrer un bénéfice en termes de survie globale mais de maintenir un bon niveau de QdV.

Par ailleurs, des recommandations ont été proposées dans certaines localisations pour inclure la QdV comme critère de jugement principal ou comme co-critère de jugement principal (Bonnetain et al., 2017; Joly et al., 2017). Elle pourrait alors présenter les avantages d'une durée d'étude plus courte et d'une capacité à évaluer l'intérêt du traitement pour le patient (Fiteni et al., 2014). De plus, elle pourrait être substitutive de la survie globale dans un contexte métastatique (Methy et al., 2010).

3.2. La place de la QdV au sein des essais cliniques en cancérologie

Depuis 1985, la FDA recommande que les évaluations de la QdV soient incluses dans tous les essais cliniques en cancérologie (Johnson & Temple, 1985).

En 1992, Osoba (1992) expose la nécessité de considérer la QdV dans les essais cliniques de phase III en cancérologie. Ces essais, destinés à comparer un nouveau traitement à un traitement standard, visent à évaluer l'efficacité pour le patient par le biais d'un critère de jugement « centrés sur le patient ». C'est pourquoi, la QdV est généralement intégrée comme critère secondaire dans les essais de phase III (Reck et al., 2018; Satoh et al., 2014; Taphoorn et al., 2005). Ainsi, lorsque le profil d'efficacité d'un nouveau traitement indique qu'il est égal au traitement standard, des informations supplémentaires sur le bénéfice en termes de QdV sont importantes.

Bien que la QdV soit systématiquement intégrée dans les essais de phase III, elle peut également être évaluée dans les essais cliniques de phase II (Brandes et al., 2016; Maio et al.,

2017; Schadendorf et al., 2016). Les essais de phase II ont pour objectif de déterminer si le traitement, son dosage et sa fréquence d'administration, ont un effet bénéfique sur la maladie. Pour ces essais, l'EORTC a adopté la politique d'inclure l'évaluation de la QdV seulement si l'essai est envisagé d'être poursuivi en phase III, dans laquelle elle est considérée comme un critère d'intérêt (Kiebert et al., 1998). Son évaluation en phase II permettrait alors de planifier sa mesure pour la phase suivante.

L'évaluation de la QdV peut également être intégrée dans les essais de phase I à titre exploratoire. Ces essais ont pour objectif d'identifier la dose maximale tolérée et la dose recommandée pour les essais de phase II. Cette dose est généralement déterminée à partir de l'évaluation des toxicités par le clinicien selon la grille *National Cancer Institute Common Terminology Criteria for Adverse Events* (NCI-CTCAE). L'adjonction de la mesure de la QdV vise à compléter les informations de l'évaluation des NCI-CTCAE. Néanmoins, Fiteni et al. (2019) ont montré dans leur revue systématique de la littérature qu'elle avait été peu étudiée dans les essais de phase I en cancérologie et qu'aucune des études la considérant, n'avait identifié la dose recommandée à l'aide des mesures de QdV. Cependant, Anota et al. (2016) ont montré que les mesures de QdV pouvaient néanmoins compléter les résultats relatifs aux toxicités et qu'elles étaient cohérentes avec la dose maximale tolérée estimée. Par ailleurs, une version des NCI-CTCAE rapportés par le patient, PRO-CTCAE, a été développée pour compléter les résultats des NCI-CTCAE rapportés par le clinicien (Basch et al., 2014).

4. Intérêt de la qualité de vie relative à la santé

L'évaluation de la QdV présente différents intérêts à être intégrée systématiquement dans les essais cliniques en cancérologie. « Ma santé 2022 » met ainsi en avant que des indicateurs, tels que la QdV, devraient toujours être mesurés (Ma Santé 2022, 2018).

Dans certaines circonstances, les traitements contre le cancer peuvent induire des effets secondaires importants, de lourdes toxicités, entraînant peu ou aucune prolongation de la survie. Ainsi, l'évaluation de la QdV a tout intérêt à être prise en compte afin de comprendre l'impact du traitement au regard de la QdV perçue par le patient et non uniquement au regard de sa tumeur. Dans d'autres circonstances, lorsque l'on s'attend à ce que les patients recevant un nouveau traitement aient une survie globale similaire à ceux recevant un traitement standard, il y a un intérêt de le montrer en termes de QdV afin de démontrer un bénéfice clinique pour le patient.

La QdV a fait la preuve de sa valeur pronostique de la survie globale dans de nombreuses études en cancérologie et dans plusieurs localisations différentes (Ediebah et al., 2018; Quinten et al., 2014). Quinten et al. (2014) ont montré dans une méta-analyse de 30 essais cliniques contrôlés randomisés sur 11 localisations cancéreuses différentes que, pour chaque localisation cancéreuse, au moins une dimension de QdV du questionnaire QLQ-C30 fournissait une information pronostique. Ainsi, les récents résultats des études ont démontré que le niveau de QdV à l'inclusion, c'est-à-dire, avant la randomisation ou l'initiation du traitement, était un facteur pronostique de la survie globale. Cependant, les dimensions de QdV donnant des informations pronostiques peuvent différer d'une étude à une autre. Ces facteurs peuvent apporter des informations complémentaires aux facteurs pronostiques établis par des mesures cliniques. En effet, les dimensions de QdV apportent des informations différentes des mesures classiques, par exemple, le bien-être, le ressenti et le fonctionnement du patient. Par ailleurs, sa valeur pronostique se révèle parfois supérieure à certaines échelles de performance évaluées par le médecin, tel l'indice de Karnofsky (Conroy et al., 2005).

Par ailleurs, le recueil électronique de la QdV en routine, de plus en plus fréquent, permet d'améliorer la communication entre le patient et le médecin, la satisfaction du patient vis-à-vis des soins, la QdV mais aussi la survie (Basch et al., 2017; Detmar et al., 2002; Velikova et al., 2004). L'évaluation de la QdV en routine permet d'obtenir des résultats de QdV en temps réel. Ces résultats en temps réel permettent ainsi d'identifier les patients présentant une

détérioration de leur QdV, et de mettre en place rapidement des mesures correctives, en adaptant par exemple leur traitement ou en les orientant vers des soins de support adaptés. Dans ce contexte, elle devient une mesure d'intérêt, permettant l'ajustement des traitements selon les scores symptomatiques de QdV.

5. Analyse longitudinale des données de QdV

Dans les essais cliniques en cancérologie, l'objectif de l'évaluation de la QdV est d'étudier l'impact du traitement sur le niveau de QdV du patient au cours du temps. Des questionnaires de QdV sont administrés aux patients tout au long de leur prise en charge, générant alors des mesures répétées pour un même patient au cours du temps. Ainsi, afin de tenir compte de la corrélation entre les mesures de QdV d'un même patient, une analyse longitudinale est requise. Cette analyse longitudinale permet d'étudier l'évolution de la QdV. De nombreuses méthodes d'analyse longitudinale des données de QdV ont été proposées. Parmi les plus utilisées figurent les modèles linéaires mixtes (Cnaan et al., 1997; Fairclough, 2010) et le temps jusqu'à détérioration d'un score de QdV (Anota et al., 2013).

L'objectif de cette partie est de présenter les principales notions liées à la méthodologie de l'analyse longitudinale des données de QdV.

5.1. La population d'analyse

Avant l'initiation d'une analyse longitudinale, il est primordial de définir la population d'analyse. En général, la population d'analyse recommandée dans les essais cliniques randomisés est celle dite en intention de traiter, correspondant à l'analyse de tous les patients randomisés, quel que soit le traitement réellement administré et quel que soit le respect des critères d'éligibilité de l'étude. La population en intention de traiter permet de pallier les deux problèmes majeurs rencontrés dans un essai clinique randomisé : la non-compliance au traitement et les données manquantes (Gupta, 2011). Ainsi, cette population permet d'interpréter les résultats de façon optimale, d'éviter des biais de sélection des patients et de pouvoir comparer les études entre elles (Bonnetain et al., 2016).

Lorsque l'on s'intéresse à l'analyse longitudinale des données de QdV, la présence d'au moins un score de QdV par patient à l'inclusion est nécessaire. Cela permet de garantir la comparabilité entre les bras de traitement et d'étudier l'évolution du niveau de QdV du patient par rapport à l'inclusion. Dans ce contexte, on parle d'une population en intention de traiter modifiée. Cette population correspond à tous les patients en intention de traiter qui ont au moins un score de QdV disponible à l'inclusion. La définition de la population en intention de traiter modifiée peut varier d'une étude à une autre. Certains auteurs définissent l'ensemble des patients en intention de traiter ayant au moins une évaluation de QdV disponible à

l'inclusion ainsi qu'au cours du suivi (Boye et al., 2016; Harbeck et al., 2016).

5.2. La différence minimale cliniquement importante

L'interprétation des scores de QdV, d'une différence de scores et leur pertinence clinique sont des enjeux majeurs de l'analyse longitudinale des données de QdV (Jaeschke et al., 1989). Osoba et al. (1998) ont souligné l'importance de déterminer une différence cliniquement significative et pas seulement statistiquement significative. En effet, une petite différence de scores en moyenne pourrait être statistiquement significative lorsque la taille d'échantillon considérée est grande, mais cette petite différence pourrait ne pas être cliniquement significative et pertinente pour le patient. Des résultats statistiquement ou cliniquement significatifs sont deux notions différentes. Ainsi, la QdV ne doit pas seulement être interprétée d'un point de vue statistique, elle doit également refléter des changements ou différences significatifs pour le patient, c'est-à-dire être interprétée au sens de la différence minimale cliniquement importante (DMCI).

La DMCI a été définie comme la plus petite différence de score de QdV que les patients perçoivent comme cliniquement importante (Jaeschke et al., 1989). Cette différence peut être soit bénéfique ou soit délétère, et entraîner le patient ou le clinicien à envisager un changement dans sa prise en charge (Schünemann & Guyatt, 2005).

Deux catégories de méthodes ont été proposées pour estimer la DMCI : les méthodes basées sur l'ancre (« *anchor-based methods* ») et les méthodes basées sur la distribution (« *distribution-based methods* ») (Crosby et al., 2003; Lydick & Epstein, 1993).

Les méthodes basées sur l'ancre utilisent un critère externe (appelé une ancre) pour déterminer les différences ou les changements dans les scores de QdV, en reliant le score de QdV étudié à un critère externe. Les ancres peuvent être cliniques (e.g. l'indice de Karnofsky ou l'évaluation des NCI-CTCAE) ou basées sur les patients (e.g. la question de transition de Jaeschke ou le score de QdV). L'une des approches basées sur l'ancre la plus utilisée pour déterminer la DMCI est la question de transition de Jaeschke (Crosby et al., 2003). Cette question est généralement formulée de la manière suivante (Jaeschke et al., 1989) :

« Au cours de ces trois derniers mois, considérez-vous que votre niveau de qualité de vie :

- est resté globalement inchangé,
- s'est détérioré : un peu, assez, beaucoup,

- s’est amélioré : un peu, assez, beaucoup. »

Les méthodes basées sur la distribution reposent uniquement sur les distributions statistiques des scores de QdV. Ces méthodes ne prennent pas en compte le point de vue des patients et des cliniciens (Musoro et al., 2018). Parmi les méthodes basées sur la distribution figurent, par exemple, l’écart-type, l’erreur standard de mesure et la taille d’effet (Copay et al., 2007).

La DMCI peut être déterminée au niveau du groupe et au niveau individuel. Au niveau du groupe, elle peut être estimée par le biais des méthodes basées sur l’ancre et sur la distribution. Au niveau individuel, la méthode basée sur l’ancre en utilisant l’approche de la courbe ROC (« *receiver operating characteristic* ») est généralement utilisée.

En ce qui concerne le questionnaire QLQ-C30, une différence de 5 à 10 points pour un score de ce questionnaire correspond au plus petit changement que les patients perçoivent comme cliniquement important (Osoba et al., 1998).

Par conséquent, déterminer ce que représente la DMCI dans les scores de QdV est utile pour les patients, les cliniciens et les chercheurs. Cela influence l’évaluation de l’efficacité des nouvelles stratégies de traitement, le calcul du nombre de sujets nécessaires pour la conception des essais cliniques lorsque la QdV est considérée comme critère ou co-critère de jugement principal (Musoro et al., 2018; Schünemann & Guyatt, 2005).

5.3. Méthodes d’analyse longitudinale les plus utilisées

5.3.1. Le modèle linéaire à effets mixtes

Le modèle linéaire à effets mixtes permet d’évaluer l’évolution des scores de QdV au cours du temps. Ce modèle est une extension du modèle linéaire prenant en compte la variabilité liée aux individus. Il se compose de deux parties, une partie à effets fixes et une partie à effets aléatoires. La partie fixe est commune à l’ensemble des individus et représente l’effet population c’est-à-dire la trajectoire moyenne du score de QdV. La partie aléatoire est propre à chaque individu. Elle permet de tenir compte de la corrélation entre les mesures de QdV et d’étudier la variabilité liée à chaque individu.

Une des modélisations possible du modèle mixte, est d’intégrer des effets fixes sur le bras de traitement, le temps et l’interaction entre le bras de traitement et le temps. L’effet bras de traitement correspond à l’écart de niveau de QdV entre les deux bras de traitement à

l'inclusion, qui est toujours présent au cours du temps. L'effet temps correspond à l'effet temps global entre les deux bras de traitement. L'effet d'interaction permet d'étudier si le niveau de QdV évolue différemment entre les deux bras de traitement. Enfin, des effets aléatoires peuvent être introduits afin de tenir compte des variations inter-individuelles notamment sur les patients et le temps.

Dans ce contexte, le modèle linéaire à effets mixtes s'écrit de la manière suivante (Anota, 2016) :

$$Y_n^{(t)} = \mathbf{c} + \alpha \times \mathbf{a}_n + \beta \times t + \gamma \times \mathbf{a}_n \times t + \delta_n + \theta \times t + \varepsilon_n^{(t)}$$

$$\varepsilon_n^{(t)} \sim \mathcal{N}(0, \sigma^2)$$

- $Y_n^{(t)}$ correspond au score du patient *n* au temps *t*,
- **c** correspond à l'intercept,
- $\mathbf{a}_n$ correspond au bras de traitement du patient *n* (cela vaut 0 lorsqu'il s'agit du bras contrôle et 1 lorsqu'il s'agit du bras expérimental),
- α correspond à l'effet fixe bras de traitement,
- β correspond à l'effet fixe temps,
- γ correspond à l'effet d'interaction entre le bras de traitement et le temps,
- δ_n correspond à l'intercept aléatoire du patient *n*,
- θ correspond aux pentes aléatoires du temps,
- $\varepsilon_n^{(t)}$ correspond à l'erreur aléatoire du patient *n* au temps *t*.

Ainsi, **c** représente le niveau initial de QdV du bras contrôle et (**c** + α) celui du bras expérimental. La pente du bras contrôle est représentée par β et celle du groupe expérimental par (β + γ). Les évolutions moyennes du score de QdV sont observées par la partie fixe. Dans la partie aléatoire, δ_n permet de considérer que les patients n'ont pas nécessairement le même niveau initial de QdV et θ que le niveau de QdV au cours du temps n'évolue pas de la même manière entre les patients.

Ces modèles requièrent de spécifier la structure de la matrice de covariance entre les mesures de QdV. Les quatre structures les plus utilisées pour l'analyse de données longitudinales sont les matrices suivantes : non structurée (« *unstructured* », UN), autorégressive d'ordre 1 à variances hétérogènes (« *heterogeneous first-order autoregressive* », ARH(1)), autorégressive

d'ordre 1 (« *first-order autoregressive* », AR(1)) et symétrie composée à variances hétérogènes (« *heterogeneous compound symmetry* », CSH). La matrice UN est utilisée lorsqu'aucune hypothèse ne peut être faite sur la structure de la matrice de covariance. Les matrices ARH(1) et AR(1) considèrent que la corrélation entre les mesures de QdV décroit au cours du temps. La structure AR(1) suppose également que les variances sont égales. Enfin, la matrice CSH suppose que les variances ne sont pas égales. Elle considère que la corrélation entre les mesures de QdV est constante au cours du temps (Blanchin et al., 2011). La structure adéquate de la matrice de covariance est sélectionnée selon le critère d'information d'Akaike.

Les modèles linéaires à effets mixtes présentent quelques limites. Ce modèle doit respecter l'hypothèse de la normalité du score de QdV, une hypothèse rarement respectée. Le choix des effets introduits dans le modèle et dans la structure de la matrice de covariance est à la main de l'utilisateur. De plus, les cliniciens sont peu familiarisés avec l'interprétation des résultats de cette méthode.

5.3.2. Le temps jusqu'à détérioration d'un score de QdV

Le temps jusqu'à détérioration d'un score de QdV est une méthode de type « temps jusqu'à événement » basée sur l'analyse de survie, où l'événement d'intérêt est la « détérioration ». Cette méthode est régulièrement utilisée pour l'analyse longitudinale des données de QdV, et plus particulièrement dans les essais cliniques de phase III (Bonnetain et al., 2016).

L'approche du temps jusqu'à détérioration d'un score de QdV nécessite une définition claire de ce qui est considéré comme une « détérioration ». Cela dépend à la fois de la localisation cancéreuse, de la situation thérapeutique, du score de référence et de la DMCI perçue par le patient qui sont utilisés pour caractériser la détérioration, ainsi que des règles de censure.

La notion de « détérioration » requiert un score de référence permettant de comparer l'évolution des scores de QdV au cours du temps comparativement à ce point de référence. Ainsi, différents scores de références ont été proposés :

- le niveau de QdV à l'inclusion du patient qui est la référence la plus utilisée (« *baseline score* »),
- le meilleur niveau de QdV déjà observé par le patient (« *best previous score* »),
- ou le score de QdV précédent la mesure actuelle du patient (« *immediately preceding score* »).

Les deux derniers scores de référence ont été proposés pour explorer l'impact potentiel de l'effet « *response shift* », reflétant le processus d'adaptation du patient à la maladie et au traitement. La notion d'effet « *response shift* » sera détaillée au paragraphe 5.5.

La « détérioration » repose sur la notion de la DMCI. Cette dernière permet de caractériser si la détérioration est cliniquement significative. Une DMCI au niveau individuel devrait être utilisée au sein de cette méthode, mais cela est rarement fait. La DMCI considérée pour le temps jusqu'à détérioration est généralement la même pour chaque dimension, mais elle peut également être différente selon les dimensions (Cocks et al., 2012) ou être différente selon le niveau de QdV du patient (e.g. une détérioration de 10% comparée au score de référence).

La détérioration pourrait être considérée comme réversible, confirmée (i.e. une détérioration maintenue pendant une période de temps définie) ou définitive (i.e. une détérioration maintenue dans le temps jusqu'à la dernière évaluation de QdV disponible de chaque patient). Des événements composites peuvent également être introduits dans la définition de l'événement, comme par exemple le décès (Bonnetain et al., 2010), la progression de la maladie (Taphoorn et al., 2018) et la détérioration d'au moins un score de QdV parmi plusieurs scores sélectionnés (Brahmer et al., 2017).

En 2013, des recommandations initiales ont proposé plusieurs définitions différentes du temps jusqu'à détérioration d'un score de QdV (Anota et al., 2013). Ces recommandations ont souligné la nécessité d'adapter la définition en fonction du cadre thérapeutique. Le *Tableau 4* ci-dessous présente les différentes définitions de l'approche du temps jusqu'à détérioration initialement proposées.

Tableau 4. Résumé des différentes définitions du temps jusqu'à détérioration (TJD) et du temps jusqu'à détérioration définitive (TJDD) d'un score de QdV d'après Anota et al. (2013).

Définitions	Score de référence			Définitive par rapport au score de référence			Décès	Patients sans baseline	Patients sans suivi
	Baseline	Best previous score	Immediately preceding score	DMCI+	DMCI-	Score qualifiant la détérioration			
TJD									
1	X								
2	X							X	X
3	X						X		
4	X						X	X	X
5		X							
6		X						X	X
7		X					X		
8		X					X	X	X
9			X						
10			X					X	X
11			X				X		
12			X				X	X	X
TJDD									
1	X			X					
2	X			X				X	X
3	X			X			X		
4	X			X			X	X	X
5	X				X				
6	X				X			X	X
7	X				X		X		
8	X				X		X	X	X
9	X					X			
10	X					X		X	X
11	X					X	X		
12	X					X	X	X	X
13		X		X					
14		X		X				X	X
15		X		X			X		
16		X		X			X	X	X
17		X			X				
18		X			X			X	X
19		X			X		X		
20		X			X		X	X	X
21		X				X			
22		X				X		X	X
23		X				X	X		
24		X				X	X	X	X
25			X	X					
26			X	X				X	X
27			X	X			X		
28			X	X			X	X	X
29			X		X				
30			X		X			X	X
31			X		X		X		
32			X		X		X	X	X
33			X			X			
34			X			X		X	X
35			X			X	X		
36			X			X	X	X	X

Une croix (X) indique la définition retenue et les événements correspondants; DMCI+, détérioration sans autre amélioration par rapport au score de référence; DMCI-, détérioration définitive si la détérioration est observée à tous les temps de mesure suivant la détérioration initiale.

En situation adjuvante, il a été proposé d'utiliser le temps jusqu'à détérioration (TJD, détérioration réversible). Le TJD a été défini comme l'intervalle de temps entre l'inclusion ou la randomisation du patient et l'apparition d'une première détérioration cliniquement significative du score de QdV par rapport au score de référence. Les patients ne présentant aucune détérioration avant leur sortie d'étude sont censurés à la dernière date de QdV disponible.

En revanche, en situation avancée ou métastatique, il a été proposé d'utiliser le temps jusqu'à détérioration définitive (TJDD), incluant ou non le décès comme événement. Le TJDD a été défini comme l'intervalle de temps entre l'inclusion ou la randomisation du patient et l'apparition d'une première détérioration cliniquement significative du score de QdV par rapport au score de référence, sans amélioration ultérieure cliniquement significative par rapport au score de référence. Les patients présentant une détérioration cliniquement significative du score de QdV, suivie de données manquantes monotones sont considérés en détérioration définitive. Les patients ne présentant aucune détérioration avant leur sortie d'étude ou ceux présentant une détérioration mais suivie d'une amélioration cliniquement significative du score de QdV par rapport au score de référence sont censurés à la date de la dernière évaluation de QdV disponible. Lorsque le décès est inclus dans la définition de l'événement on parle de survie sans détérioration (QFS, « *health-related quality of life deterioration-free survival* »).

De nouvelles recommandations autour de la méthode du temps jusqu'à détérioration sont en cours de développement par le *Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data* (SISAQoL) consortium. En 2020, le groupe SISAQoL a proposé une standardisation de la définition de la détérioration (Coens et al., 2020). Deux définitions découlent de ces recommandations :

- la détérioration transitoire, qui a été définie comme l'intervalle de temps entre l'inclusion ou la randomisation du patient et l'apparition d'une première détérioration cliniquement significative du score de QdV par rapport au score à l'inclusion. Après la première détérioration, au moins un score de suivi doit être supérieur ou au moins être égal au score à l'inclusion.

- et la détérioration définitive, qui a été définie comme l'intervalle de temps entre l'inclusion ou la randomisation du patient et l'apparition d'une première détérioration cliniquement significative du score de QdV par rapport au score à l'inclusion, tout en

maintenant cette détérioration cliniquement significative pour tous les scores suivants (i.e. la détérioration est observée pour tous les scores suivants la première détérioration).

Cependant, ces recommandations ne tiennent pas compte de la localisation cancéreuse ni de la situation thérapeutique. Le score de référence se limite au score à l'inclusion.

Étant donné que l'approche du temps jusqu'à détérioration d'un score de QdV s'apparente aux analyses de survie, son estimation peut être calculée à l'aide de la méthode d'estimation de Kaplan-Meier (Goel et al., 2010). L'estimateur de Kaplan-Meier permet d'estimer la fonction de survie et se base sur l'idée qu'être en vie jusqu'à un instant *t*, c'est être en vie à l'instant *t* sachant que l'on était encore en vie à l'instant ($t - 1$). Dans cette définition, l'événement est le décès. Dans les analyses du temps jusqu'à détérioration, l'événement est la détérioration du score de QdV. Ainsi, l'estimateur de Kaplan-Meier est donné par la formule suivante :

$$\hat{S}(t) = \prod_{t_i \leq t} \frac{n_i - d_i}{n_i}$$

Où $n_i = n_{i-1} - d_{i-1} - c_{i-1}$

- n_i correspond au nombre de patients à risque juste avant le temps *i*, i.e. le nombre de patients encore dans l'étude et qui ne présente pas de détérioration avant le temps ($i - 1$),
- d_i correspond au nombre de patients qui se sont détériorés au temps *i*,
- c_i correspond au nombre de patients censurés au temps *i*, i.e. le nombre de patients ne présentant aucune détérioration avant leur sortie d'étude au temps *i*.

Par la suite, le temps jusqu'à détérioration peut être comparé par bras de traitement dans le cas d'essais cliniques randomisés utilisant le test du log-rank et des analyses de Cox univariées pour produire des rapports de risques, nommé hazard ratio (Cox, 1972). Un modèle de Cox multivarié peut être réalisé afin d'étudier les facteurs influençant potentiellement le temps jusqu'à détérioration. Le modèle de Cox s'écrit de la manière suivante :

$$\lambda(t, X_1, \dots, X_n) = \lambda_0(t) \exp\left(\sum_{i=1}^{n} \beta_i X_i\right)$$

- $\lambda(t,X_1,\ldots,X_n)$ correspond au risque instantané de détérioration au temps t sachant que le patient n'est pas détérioré juste avant le temps t,
- $\lambda_0(t)$ correspond au risque de base, i.e. le risque instantané de détérioration lorsque toutes les covariables sont nulles,
- $X_1,\ldots,X_n$ correspond aux covariables,
- β_i correspond à l'effet des covariables sur le risque instantané.

Ce modèle repose sur le respect de l'hypothèse de proportionnalité des risques. Si cette hypothèse n'est pas respectée, la méthode des « *restricted mean survival time* » (RMST) peut être utilisée en complément du hazard ratio afin d'assurer la robustesse du modèle (Royston & Parmar, 2011, 2013). Le RMST est défini comme l'aire sous la courbe de la fonction de survie jusqu'à un temps t^*. Ainsi, dans un essai clinique à deux bras avec des fonctions de survie $S_0(t)$ et $S_1(t)$ dans le bras de contrôle et le bras expérimental, respectivement, la différence Δ de RMST entre les deux bras de traitement est :

$$\Delta = \int_0^{t^*} [S_1(t) - S_0(t)]\, dt$$

Où $t^* > 0$

Ainsi, la différence Δ de RMST mesure le gain ou la perte d'espérance de vie au temps t^* associé au bras expérimental. Une différence Δ de RMST supérieure à zéro favorisera le bras expérimental (Trinquart et al., 2016).

Cette méthode d'analyse longitudinale est attractive pour de nombreux cliniciens puisqu'elle propose des résultats facilement compréhensibles et interprétables de par la méthodologie utilisée.

5.4. Les données manquantes

L'occurrence de données manquantes est un enjeu majeur pour l'analyse longitudinale des données de QdV (Fairclough et al., 1998). Les données manquantes sont inévitables dans la majorité des essais cliniques en cancérologie. En effet, les questionnaires de QdV sont rarement complétés intégralement à chaque temps de mesure prévu. Les données manquantes

se produisent lorsque certains items ne sont pas renseignés ou lorsque le questionnaire dans son ensemble n'est pas rempli. Certaines peuvent être étroitement liées au niveau de santé et de QdV des patients, résultant en des estimations erronées de l'effet du traitement (Sloan et al., 2007). Elles peuvent être dues à plusieurs raisons, par exemple, le patient manque un rendez-vous planifié, le personnel oublie d'administrer le questionnaire, le patient refuse de remplir le questionnaire, le patient déclare être trop malade pour répondre au questionnaire ou le patient décède (Fairclough, 2010). Ainsi, ignorer la problématique des données manquantes pourrait introduire des biais dans les résultats des analyses. Il est donc primordial de trouver des solutions appropriées pour traiter ce phénomène.

5.4.1. Les différents types de données manquantes

Dans le cadre d'une évaluation de la QdV à plusieurs temps de mesure, différents types de données manquantes peuvent être observés : intermittentes, monotones et mixtes (Fielding et al., 2012).

Intermittentes. La donnée de QdV est manquante à un ou plusieurs temps de mesure, mais cette donnée est renseignée à un temps de mesure ultérieur. Une donnée manquante intermittente peut correspondre à un patient qui a omis de répondre à un item au sein de son questionnaire de QdV ou d'une dimension de QdV. Elle peut également correspondre à un patient qui ne remplit pas son questionnaire de QdV à un temps de mesure, mais qui en remplit un par la suite lors d'un suivi ultérieur.

Monotones. Les données de QdV sont manquantes à partir d'un certain temps de mesure et pour tous les temps de mesure suivants. Les données manquantes monotones peuvent correspondre à un patient qui ne remplit plus ses questionnaires de QdV à un temps de mesure ainsi que pour toutes les évaluations suivantes. Dans les essais cliniques en cancérologie, ces données manquantes correspondent principalement à la sortie d'étude du patient, dues par exemple à la progression de la maladie ou au décès. Dans ce cas, les données manquantes monotones sont prévues et informatives du niveau de QdV du patient.

Mixtes. Les données manquantes mixtes correspondent à une période de données manquantes intermittentes, suivie de données manquantes monotones.

5.4.2. Classification des données manquantes

Little et Rubin (1987) ont proposé une classification des données manquantes : les données manquantes complètement aléatoires, aléatoires et non aléatoires. Ces différents mécanismes décrivent les raisons pour lesquelles des données manquantes se sont produites (Fielding et al., 2012).

Manquantes complètement aléatoires. Les données manquantes complètement aléatoires (« *missing completely at random* », MCAR) sont indépendantes des données observées, i.e. les données cliniques et socio-démographiques ainsi que du niveau de QdV observé, et non observées, i.e. le niveau de QdV manquant du patient. Celles-ci sont sans rapport avec le passé, le présent et le futur. Par exemple, si un patient déménage ou si le personnel hospitalier (e.g. l'attaché de recherche clinique) oublie d'administrer un questionnaire, alors les données manquantes observées sont complètement dues au hasard.

Manquantes aléatoirement. Les données manquantes aléatoires (« *missing at random* », MAR) sont dépendantes des données observées et indépendantes des données non observées. Par exemple, si un patient ne répond pas à certaines questions pouvant être gênantes (e.g. questions relatives à la sexualité), alors les données manquantes observées sont aléatoires. Ainsi, celles-ci pourraient s'expliquer par les données cliniques et socio-démographiques recueillies à l'inclusion telles que l'âge ou le sexe. Ces données manquantes pourraient également être expliquées par les scores de QdV observés précédant la donnée manquante. Dans la plupart des cas, il s'agirait de l'évaluation à l'inclusion.

Manquantes non aléatoirement. Les données manquantes non aléatoires (« *missing not at random* », MNAR) sont indépendantes des données observées et dépendantes des données non observées, c'est-à-dire le niveau de QdV manquant du patient. Par exemple, si un patient est dans l'incapacité de remplir le questionnaire due à la présence de symptômes tels que la fatigue, les nausées et vomissements, ou la progression tumorale, alors les données manquantes ne sont pas dues au hasard. Ces données manquantes peuvent être probablement liées au niveau de QdV du patient.

Le terme de données manquantes ignorables (i.e. non informatives) est souvent assimilé aux données MCAR et MAR, et à l'inverse, le terme de données manquantes non ignorables (i.e.

informatives) aux données MNAR.

5.4.3. Détermination du mécanisme des données manquantes

La détermination du mécanisme des données manquantes est un processus complexe. Elle permet d'aider à identifier la méthode d'analyse la plus appropriée à appliquer aux données en présence de données manquantes.

L'identification d'un mécanisme MCAR peut être effectuée de diverses façons. Little (1988) a développé un test basé sur la comparaison des moyennes du score observées pour chaque schéma (dit pattern) de données manquantes. Les données manquantes seront considérées comme MCAR si les moyennes des données observées sont identiques pour chaque pattern de données manquantes. Dans le cas contraire, un mécanisme MAR sera mis en évidence si les moyennes sont différentes. Fairclough (2010) propose que les différences entre les mécanismes MCAR et MAR peuvent être évaluées en examinant l'association des scores de QdV observés et les données manquantes, par le biais d'une régression logistique. L'identification d'un mécanisme MCAR ou MAR peut également être mise en évidence en comparant les patients qui ont au moins un score de QdV disponible à l'inclusion aux autres patients. Cette comparaison est réalisée selon les données cliniques et socio-démographiques de chaque patient, recueillies à l'inclusion, en utilisant un test de Student ou le test non-paramétrique de Mann-Whitney pour les variables continues, et un test du χ^2 ou un test exact de Fisher pour les variables qualitatives. Ces analyses permettent de déterminer si les données manquantes à l'inclusion dépendent ou non des caractéristiques du patient.

L'identification d'un mécanisme MNAR reste complexe, puisque les données dont nous avons besoin pour tester l'hypothèse sont manquantes, i.e. le niveau de QdV manquant du patient. Cependant, il est possible de recueillir des informations qui peuvent mettre en évidence un mécanisme MNAR. En effet, lorsque d'autres mesures en lien avec la maladie ou le traitement, telles que la progression de la maladie, sont recueillies et fortement associées à la QdV, il est possible que les données soient MNAR (Fairclough, 2010). Cette identification de mécanisme pourrait être exploitée, par exemple, en comparant la survie globale des patients ayant au moins un score de QdV disponible à l'inclusion aux autres patients ou par le biais du recueil des raisons de non-remplissage des questionnaires.

5.4.4. Gestion des données manquantes

Ces dernières années, différentes méthodes d'analyses dépendant du mécanisme de données manquantes ont été proposées et testées dans des applications cliniques (Sloan et al., 2007).

- L'analyse des cas complets :

L'option la plus simple mais généralement la moins souhaitable pour traiter les données manquantes consiste à les ignorer complètement et à entreprendre une analyse des cas complets. Elle consiste à limiter l'analyse aux seuls patients n'ayant aucune donnée manquante. Cette méthode d'analyse supprime potentiellement un grand nombre de patients de l'analyse et est susceptible de fournir des résultats biaisés, à moins que le mécanisme soit MCAR et que la proportion de données manquantes soit faible (Fielding et al., 2012). La seule conséquence lorsque les données sont MCAR est une perte de puissance due au nombre réduit de patients analysés. L'analyse des cas complets produira des résultats biaisés si les données sont MAR ou MNAR (Fielding et al., 2016).

Une deuxième option souvent envisagée pour traiter les données manquantes est l'utilisation d'imputation. L'imputation est un processus de remplacement des données manquantes avec des valeurs substituées. La donnée manquante peut être remplacée par une seule valeur (imputation simple) ou par plusieurs valeurs (imputation multiple). Lorsque l'imputation est réalisée, un jeu de données complet est obtenu.

- L'imputation simple :

L'imputation simple permet de remplacer chaque valeur manquante par une valeur unique. Diverses méthodes d'imputation simple ont été proposées.

- Imputation simple par la moyenne :

La méthode d'imputation simple par la moyenne est couramment utilisée en QdV et peut se décomposer en différentes méthodes : les méthodes d'imputation simple par le score moyen personnel (« *personal mean score* »), par la moyenne par item (« *item mean score* ») et par la

moyenne par item corrigée (« *corrected item mean* ») (Huisman, 2000).

La méthode d'imputation simple « *personal mean score* » consiste à imputer la donnée manquante par le score moyen de l'individu selon les réponses observées aux autres items. Cette méthode considère que l'item manquant ne diffère pas en moyenne des autres items renseignés pour la dimension concernée, et ainsi que l'item manquant n'apportera pas d'information supplémentaire par rapport aux items renseignés. Ceci correspond aux recommandations proposées dans plusieurs manuels de scoring, à condition qu'au moins 50% des items de la dimension considérée soient renseignés.

La méthode d'imputation simple « *item mean score* » consiste à imputer l'item manquant par la moyenne des observations des patients pour cet item.

Ainsi, les valeurs imputées selon ces deux méthodes sont centrées par rapport aux scores observés, soulignant l'hypothèse que les valeurs manquantes sont MCAR (Fairclough, 2010).

La méthode « *personal mean score* » prend en compte la capacité du patient, i.e. son niveau de QdV, alors que la méthode « *item mean score* » ne prend en compte que la difficulté de l'item, i.e. plus l'item est difficile, moins il sera probable de répondre à une modalité élevée. La troisième méthode « *corrected item mean* » est une combinaison des deux précédentes méthodes : la moyenne par item est pondérée par le score moyen personnel de l'individu (Hardouin et al., 2011).

- Imputation simple par le minimum ou le maximum :

La méthode d'imputation simple par le minimum ou le maximum consiste à imputer la donnée manquante par un score de QdV élevée ou faible. Cette méthode est souvent utilisée lorsque les données de QdV manquantes sont le résultat d'un événement indésirable grave tel que le décès (Fairclough, 2010). Ainsi, dans le cas de données manquantes dues au décès, le score de QdV pourrait être imputé par une valeur égale à zéro. Cependant, cette méthode reste controversée et n'est pas recommandée à ce jour (Revicki et al., 2001).

- Imputation simple « *last observation carried forward* » :

La méthode d'imputation simple « *last observation carried forward* » consiste à imputer la donnée manquante par la dernière valeur disponible observée. Cette méthode fait l'hypothèse que le résultat reste inchangé dans le temps, c'est-à-dire que le niveau de QdV du patient reste

constant au cours du temps depuis la dernière valeur disponible, ce qui semble peu probable (Fielding et al., 2008).

- Imputation simple par un modèle de régression :

L'imputation simple par un modèle de régression consiste à imputer la donnée manquante par une donnée prédite obtenue à partir d'une régression de la variable de l'item manquant sur les autres items de la dimension concernée. Les données utilisées pour ce modèle sont issues de tous les patients avec des informations complètes sur tous les items de la dimension concernée.

- Imputation simple « *hot-deck* » :

L'imputation simple « *hot-deck* » sélectionne au hasard, parmi les patients avec des données de QdV observées, le score de QdV de l'un d'entre eux et l'impute au patient ayant ce score manquant. La sélection du score peut se faire par le biais d'un patient possédant des caractéristiques proches du patient ayant la donnée manquante.

Les méthodes d'imputation simple sont susceptibles de fournir des résultats biaisés. En effet, la majorité des méthodes supposent que les données manquantes sont MCAR (Fielding et al., 2012). De plus, l'imputation simple s'est avérée inadéquate en présence d'un mécanisme MNAR (Fielding, 2008).

- L'imputation multiple :

L'imputation multiple est un processus qui consiste à imputer plusieurs valeurs à une donnée manquante. Les techniques d'imputation multiple permettent de tenir compte de l'incertitude de la vraie valeur de la donnée manquante, en créant plusieurs jeux de données. Rubin (2004) préconise de réaliser entre trois et cinq imputations par donnée manquante. La méthode d'analyse choisie est effectuée sur chaque jeu de données et les résultats sont combinés pour obtenir un seul résultat global. Plusieurs approches peuvent être utilisées, notamment la méthode de régression, la méthode du score de propension ou encore la méthode de Monte-Carlo par Chaînes de Markov. Théoriquement, les imputations multiples peuvent traiter les données manquantes lorsque le mécanisme est MAR ou MNAR. Lorsque le mécanisme est

MNAR, le processus de sortie d'étude peut être modélisé et incorporé dans la procédure d'imputation multiple. Cependant, comme ce modèle ne peut pas être vérifié car les données requises sont manquantes, la plupart des méthodes d'imputation multiple requièrent l'hypothèse que les données manquantes sont de mécanisme MAR (Fielding et al., 2008). Le choix de la méthode dépend du type de variable et du type de données manquantes. L'imputation multiple reste complexe mais les récents développements sur logiciel la rendent plus facilement accessible.

Enfin, il est possible de traiter la problématique des données manquantes par le biais de méthodes d'analyse longitudinale. Lorsqu'elles ont un mécanisme MAR, la modélisation de modèles linéaire à effets mixtes est adaptée. Des procédures plus complexes qui permettent de tenir compte d'un mécanisme MNAR existent, telles que les modèles de type « *pattern mixture* » (Fielding et al., 2012). Dans l'approche du temps jusqu'à détérioration d'un score de QdV, les données manquantes intermittentes ou monotones sont généralement ignorées, considérant que le niveau de QdV du patient reste constant depuis la dernière mesure de QdV disponible. De plus, les données manquantes dues au décès peuvent être prises en compte lorsque le décès est intégré dans la définition de l'événement. Des analyses de sensibilité pourraient également être effectuées en considérant que la donnée manquante est informative et ainsi considérer le patient en détérioration à la date théorique d'évaluation de la QdV.

5.5. L'effet « *response shift* »

La QdV étant un concept subjectif et dynamique, les attentes et espérances de santé peuvent évoluer au cours du temps, dues au processus d'adaptation des patients à la maladie et au traitement. Ce processus se traduit par l'effet « *response shift* » ou « changement de réponse ». Sprangers et Schwartz (1999) ont défini la « *response shift* » comme un ensemble de trois composantes correspondant à :

- un changement dans les références internes du patient (le recalibrage, « *recalibration* »),
- un changement dans l'importance des différentes dimensions de QdV (la redéfinition des priorités, « *reprioritization* »),
- un changement dans la définition du concept de la QdV (la reconceptualisation, « *reconceptualization* »).

La « *recalibration* » représente un changement de jugement du patient au cours du temps. Pour cela, imaginons que nous demandons à un patient d'évaluer son niveau de QdV au moment du diagnostic de sa maladie puis après l'initiation de son traitement, d'évaluer son niveau de QdV actuel et de le réévaluer au moment du diagnostic de la maladie. Le patient pourrait revoir son jugement à la hausse ou à la baisse sur son niveau initial de QdV, ainsi il pourrait recalibrer son niveau de QdV. Par exemple, un patient donne un score de QdV au moment du diagnostic de 20 sur une échelle qui varie de 0 (pire) à 100 (meilleur). Puis, après l'initiation de son traitement, ce même patient donne encore un score de QdV de 20 et estime que son score de QdV au moment du diagnostic était de 80. Ainsi, ce patient a revu son jugement initial à la hausse et a recalibré son niveau de QdV au moment du diagnostic. La « *reprioritization* » représente un changement dans l'ordre d'importance des différentes dimensions de QdV d'un individu. Par exemple, avant le diagnostic d'une maladie, un patient peut accorder plus d'importance à la composante physique. Après diagnostic, ce même patient pourrait revoir ses priorités et se tourner vers une dimension spirituelle ou familiale. Enfin, ce changement d'ordre d'importance des différentes dimensions de QdV d'un individu pourrait entraîner un changement chez cet individu de sa définition du concept de la QdV, la « *reconceptualization* ».

Plusieurs méthodes ont été proposées pour mettre en évidence l'occurrence d'un effet « *response shift* » (Schwartz & Sprangers, 1999), comme par exemple, la méthode « *then-test* » (Sprangers et al., 1999) ou les modèles à équations structurelles (Oort, 2005).

Ainsi, identifier l'occurrence d'un effet « *response shift* » dans l'analyse longitudinale des données de QdV permettrait de tenir compte d'une sous-estimation ou surestimation de l'effet du traitement. En effet, les patients ne pouvant pas avoir recours à ce processus d'adaptation pourraient avoir tendance à se détériorer plus rapidement et diminuer leur survie globale. Le fait de détecter ces patients permettrait, par exemple, en pratique clinique de leur proposer des soins de support.

6. Challenge de l'analyse longitudinale des données de QdV

La QdV est un critère de jugement important dans les essais cliniques en cancérologie pour évaluer le bénéfice clinique de nouvelles stratégies de traitement pour le patient (Osoba, 2011; Secord et al., 2015). Cependant, l'analyse longitudinale des données de QdV reste complexe (Bonnetain et al., 2016), puisque les diverses façons de les analyser et de les interpréter, rendent difficile la comparaison des résultats entre les différents essais cliniques, entraînant la dévalorisation de l'utilisation de données de QdV (Bottomley et al., 2016; Pe et al., 2018). Dans les années à venir, il sera primordial que des recommandations sur la standardisation des définitions des critères de jugements, des méthodes d'analyses, de l'interprétation et de la présentation des données de QdV soient faites afin de pouvoir renforcer l'utilisation de telles données dans les essais cliniques en cancérologie. La résolution de cette problématique est désormais ancrée dans l'esprit des chercheurs. Des recommandations sur la standardisation de l'analyse et de l'interprétation des données de type PROs sont en cours de développement, notamment, par le groupe SISAQoL. De plus, de récentes recommandations ont été proposées pour faciliter la communication des résultats des données de PROs, grâce au développement de la checklist « *Consolidated Standards of Reporting Trials – PRO* » (Calvert et al., 2013). Par ailleurs, le « *Standard Protocol Items: Recommendations for Interventional Trials – PRO* » a récemment été développé fournissant des recommandations autour des éléments devant être traités et inclus dans les protocoles des essais cliniques lorsque les mesures de PROs y sont intégrées (Calvert et al., 2018).

L'objectif de cette partie est de présenter les différentes problématiques de l'analyse longitudinale des données de QdV, qui peuvent également s'étendre aux données de PROs.

6.1. La différence minimale cliniquement importante

La détermination d'une DMCI robuste et fiable reste difficile et constitue un enjeu majeur pour l'analyse longitudinale des données de QdV. En effet, cette détermination dépend de la taille de l'échantillon pouvant ne pas être suffisante pour obtenir une DMCI fiable.

Plusieurs approches pour estimer la DMCI d'un questionnaire de QdV devraient être réalisées. Cependant, l'utilisation de multiples approches peuvent produire différentes valeurs entraînant des décisions sur la sélection d'une valeur unique ou d'une gamme de valeurs.

Revicki et al. (2008) ont recommandé que l'estimation de la DMCI devait être principalement des ancres cliniques et basées sur le patient, qui sont corrélées avec les dimensions du questionnaire. L'ancre utilisée doit être pertinente dans le contexte de la maladie et avoir une relation conceptuelle entre elle et le patient (Musoro et al., 2018). Le choix des ancres devrait être réalisé par les chercheurs et les cliniciens, afin que son interprétation soit claire d'un point de vue médical. Ainsi, les méthodes basées sur la distribution seront utilisées comme preuves pour appuyer et aider à l'interprétation des estimations obtenues par les méthodes basées sur l'ancre (Revicki et al., 2008). Ces méthodes pourront également être utilisées lorsqu'aucune estimation basée sur l'ancre n'est disponible. Les méthodes basées sur l'ancre et sur la distribution restent les méthodes les plus couramment utilisées pour déterminer la DMCI.

L'estimation d'une DMCI peut varier selon les dimensions, les questionnaires, les populations, le contexte clinique de l'étude et les méthodes appliquées. Par exemple, l'utilisation de différentes ancres cliniques ou basées sur les patients peut conduire à des estimations de DMCI différentes (Crosby et al., 2003). L'existence d'une unique DMCI à appliquer à chaque étude quels que soient le contexte ou les questionnaires serait insatisfaisante. Cependant, les différentes DMCI existantes peuvent complexifier le message lorsque les données de QdV sont considérées comme critère de jugement principal. En effet, si trois dimensions sont ciblées avec l'utilisation de trois DMCI différentes alors trois calculs de nombre de sujets nécessaires doivent être réalisés.

A ce jour, l'interprétation des résultats au regard de la DMCI reste encore peu utilisée (Fiteni et al., 2016). Malgré des nouveaux seuils récemment proposés, les DMCI retenues dans les études longitudinales pour les questionnaires de l'EORTC sont généralement de 5 ou 10 points. De plus, dans certaines situations une DMCI au niveau individuel devrait être utilisée, comme par exemple avec l'approche du temps jusqu'à détérioration d'un score de QdV, mais cela est rarement fait.

Ainsi, il est essentiel de conduire des études sur la détermination de la DMCI pour acquérir une uniformité dans les estimations et gagner en confiance quant à l'interprétation et la comparaison des résultats.

6.2. Méthodes d'analyse longitudinale

Le développement de guidelines sur les méthodes d'analyse longitudinale est nécessaire afin d'améliorer l'analyse et l'interprétation des données de QdV. Plusieurs revues systématiques de la littérature ont mis en évidence une absence de consensus sur la façon d'analyser et d'interpréter les données de QdV (Fiteni et al., 2016; Hamel et al., 2017; Pe et al., 2018). Par conséquent, il est difficile de tirer des conclusions solides et de comparer les résultats entre différents essais cliniques, en raison d'absence de consensus. Les nombreuses manières possibles d'analyser les données de QdV peuvent conduire à des résultats contradictoires et ainsi des décisions différentes. En effet, en l'absence de consensus le chercheur choisit lui-même la méthode d'analyse à appliquer. Dans certains cas, le choix de cette méthode peut être motivé par sa familiarité auprès des cliniciens. Bottomley et al. (2016) recommandent que le choix de la méthode statistique soit une combinaison d'une modélisation statistique robuste, prédéfinie, complétée par des présentations de résultats facilement interprétables par les cliniciens. Cependant, il n'existe pas une méthode d'analyse unique pouvant traiter tous les problèmes cliniques, de conceptions des essais et d'analyse, liés aux données de QdV (Coens et al., 2020). Le groupe SISAQoL a élaboré une liste de critères statistiques qui a été diffusée aux membres de leur groupe de recherche, dans laquelle ils ont évalué chaque critère proposé comme essentiel, souhaitable ou non essentiel pour l'analyse des données de QdV dans les essais cliniques en cancérologie. Les résultats de cette enquête ont été récemment publiés et feront l'objet de travaux supplémentaires.

Au sein même des méthodes d'analyse longitudinale des recommandations devraient être élaborées. Par exemple, malgré les recommandations initiales de la méthode du temps jusqu'à détérioration d'un score de QdV (Anota et al., 2013), d'autres définitions de détérioration ont été utilisées dans les essais cliniques randomisés en cancérologie (Cortes et al., 2015; Verma et al., 2018) qui peuvent impacter ou limiter la comparaison des résultats entre les différents essais cliniques, comme d'autres critères de jugement du type « temps jusqu'à événement » (Mathoulin-Pelissier et al., 2008). La méthode du temps jusqu'à détérioration d'un score de QdV dépend de plusieurs paramètres. Ces différents paramètres rendent cette méthode très flexible pouvant s'adapter à diverses situations, mais aussi très problématique. En effet, cette méthode est utilisée de manière non standardisée due à la variabilité des définitions existantes et des événements composites potentiellement inclus dans la définition de l'événement. Les événements associés à la définition du temps jusqu'à détérioration devraient être choisis

judicieusement et être justifiés. Par exemple, inclure la progression de la maladie en tant qu'événement composite à la détérioration pourrait être controversé, car la progression de la maladie est un critère de jugement « centré sur la tumeur » et non « centré sur le patient ». Cependant, la QdV est généralement évaluée jusqu'à la progression de la maladie dans les essais cliniques randomisés. Dans une telle situation, ignorer la progression de la maladie entraînerait une censure informative. Ainsi, il semble pertinent de tenir compte de la progression en événement composite puisque la progression de la maladie pourrait correspondre potentiellement à une détérioration de la QdV. Chaque événement composite devrait avoir une importance similaire pour le patient et ainsi être cliniquement pertinent (Montori et al., 2005) mais également avoir un effet du traitement homogène (Bonnetain et al., 2016; Freemantle et al., 2003). Par ailleurs, la méthode du temps jusqu'à détérioration d'un score de QdV est influencée par les temps de mesure. Cette méthode ne doit être appliquée qu'avec un nombre raisonnable de temps de mesure (Bonnetain et al., 2016). Le groupe EORTC recommande au moins trois évaluations de la QdV dans les essais cliniques (Kiebert et al., 1998). Afin de pouvoir appliquer cette méthode, trois temps de mesure semblent être le minimum d'évaluations à avoir. La fréquence et l'intervalle entre deux évaluations consécutives peuvent varier d'un essai clinique randomisé à un autre. Ces temps de mesure dépendent de l'évolution de la maladie et du changement attendu de la QdV. La date de l'évaluation de la QdV est prévue au préalable cependant le véritable temps de détérioration du score de QdV se situe quelque part dans l'intervalle de temps entre les deux évaluations. Ainsi, ne pas considérer cet intervalle de temps entre les deux évaluations entraîne une surestimation de la médiane de détérioration du score de QdV. Il s'agit d'un problème courant pour les critères de jugement du type « temps jusqu'à événement » (Panageas et al., 2007), et la censure par intervalle pourrait être une solution envisageable. Le groupe SISAQoL est en cours d'élaboration de recommandations tenant compte de toutes les problématiques relatives à cette méthode.

Enfin, la nature multidimensionnelle de la QdV combinée à des mesures répétées et l'occurrence de données manquantes peuvent potentiellement biaiser les résultats à cause de la multiplicité des tests. La meilleure façon de traiter la multiplicité des tests est de définir au préalable dans le protocole ou dans le plan d'analyse statistique *a priori*, les dimensions de QdV que le traitement devrait affecter et celles qui ne devraient pas être affectées (Bonnetain et al., 2016), ainsi que la méthode d'ajustement. Il semble peu probable qu'un traitement affecte uniformément une maladie dans toutes les dimensions de QdV (Sloan et al., 2007). Il

est primordial de spécifier au préalable dans le protocole les dimensions cibles ainsi que la méthode d'ajustement.

6.3. Les données manquantes

L'occurrence des données manquantes peut avoir des impacts majeurs sur l'analyse longitudinale des données de QdV. Au minimum, les données manquantes réduisent la taille de l'échantillon et la puissance statistique permettant de détecter un effet du traitement. Au pire, les données manquantes biaisent les résultats et peuvent fausser les effets du traitement (Bonnetain et al., 2016; Fielding et al., 2016; Sloan et al., 2007). Ainsi, ignorer la présence des données manquantes pourrait conduire à une conclusion erronée, par exemple, à l'adoption de la mauvaise intervention clinique. Cette conclusion pourrait donc impacter la pratique clinique (Fielding et al., 2012). En raison des impacts potentiels des données manquantes, la détermination du mécanisme des données manquantes est d'une importance capitale et doit orienter vers la méthode d'analyse longitudinale la plus appropriée.

En effet, selon le mécanisme de données manquantes identifié, l'impact sur l'analyse longitudinale est différent. Les données manquantes MCAR et MAR sont non informatives puisqu'elles sont indépendantes du niveau de QdV manquant du patient. Elles n'entraîneront pas de biais dans l'analyse longitudinale si elles sont correctement considérées. En effet, les données manquantes de type MAR n'entraîneront pas de biais uniquement si l'on prend en compte les données observées dépendantes des données manquantes. Conditionnellement à ces données manquantes, elles pourront être considérées comme MCAR. En revanche, les données manquantes MNAR sont informatives du niveau de QdV du patient et donc de son état de santé. Celles-ci peuvent entraîner un biais dans l'analyse longitudinale si elles ne sont pas correctement prises en compte dans l'analyse. En effet, si la population d'analyse est celle des patients qui ont complété le questionnaire de QdV à l'inclusion et que ces patients présentent un meilleur niveau de QdV que ceux qui n'ont pas complété ce questionnaire, les analyses seront effectuées sur une sous-population de patients présentant un bon niveau de QdV général. Ainsi, les résultats ne refléteront pas l'ensemble de la population initiale.

La meilleure solution au regard des données manquantes serait d'avoir des données complètes. Cette solution semble être irréaliste mais peut être néanmoins améliorée par de simples ajustements méthodologiques. Sloan et al. (2007) ont proposé une liste de points clés permettant de minimiser l'occurrence des données manquantes. Celle-ci conseille de :

1) Traiter les données de QdV comme n'importe quel autre critère de jugement des essais cliniques. Lorsque la QdV est considérée comme critère de jugement dans l'essai clinique, les mesures de QdV doivent être intégrées dans la collecte de données et le protocole. Celles-ci ne doivent pas être considérées comme une charge supplémentaire mais plutôt comme un élément essentiel des objectifs scientifiques de l'étude.

2) Identifier le personnel clé pour superviser et coordonner les aspects de la QdV et servir comme expert pour le feedback de l'étude.

3) S'assurer que les mesures de QdV sont administrées de manière standardisée dans tous les sites concernés par l'étude.

4) Tenir compte du format de présentation du questionnaire et de la charge des patients à répondre à plusieurs questionnaires.

Un autre point clé qui permettrait d'améliorer la problématique des données manquantes serait de récolter les raisons de non-remplissage des questionnaires dans les essais cliniques, comme le fait systématiquement le groupe EORTC (Kiebert et al., 1998). En effet, recueillir ces informations permettrait également d'aider à la détermination du mécanisme de données manquantes.

Désormais, de nombreux questionnaires de QdV validés tiennent compte des items manquants dans leurs algorithmes de calcul des scores, permettant de pouvoir réduire le pourcentage de données manquantes.

Fairclough (1998) a indiqué que lorsque le pourcentage de données manquantes est supérieur à 50% pour un critère de jugement, alors aucune analyse ne devrait être effectuée puisque les analyses seraient basées sur autant de données présentes que manquantes. Lorsque le pourcentage de données manquantes est entre 20% et 50%, les analyses peuvent être compromises mais quand même réalisées si le mécanisme des données manquantes est déterminé et accompagné par des analyses de sensibilité approfondies. Lorsque le pourcentage de données manquantes est inférieur à 20%, celles-ci n'auront pas d'impact majeur sur les résultats mais des analyses de sensibilité sont recommandées. Enfin, lorsque le pourcentage de données manquantes est inférieur à 10%, un biais potentiel dans les résultats est faible et une approche d'imputation simple semblerait suffisante.

Les approches statistiques pour traiter les données manquantes devraient être explicitement

définies a priori dans le protocole, en tant qu'analyses de sensibilité. L'intérêt de spécifier au préalable les méthodes choisies est d'éviter toute suspicion ultérieure quant à la méthode retenue et d'éviter un choix de complaisance (Fielding et al., 2012). Sloan et al. (2007) recommandent jusqu'à trois approches d'imputation différentes en analyse de sensibilité. Ces différentes analyses de sensibilité permettraient de vérifier la cohérence avec les résultats de l'analyse principale et ainsi soutenir la robustesse des résultats.

6.4. Les essais cliniques en ouvert

La méthodologie autour de la QdV reste un enjeu majeur, pour ces données dites subjectives. En 2009, la FDA a publié ses premières directives sur les PROs. Elle estime que le résumé des caractéristiques des produits, fixé lors de l'autorisation de mise sur le marché, relatif aux PROs dans les essais en ouvert, est rarement suffisant et peut être problématique (FDA, 2009; Roydhouse et al., 2019). En effet, lorsque l'essai est en ouvert, les patients connaissent le traitement attribué, ce qui pourrait influencer leurs réponses aux questionnaires et biaiser les résultats. Les patients randomisés dans le bras contrôle pourraient être déçus et ainsi sous-estimer leur niveau de QdV et surestimer leurs symptômes. Inversement, les patients randomisés dans le bras expérimental pourraient être satisfaits et ainsi surestimer leur niveau de QdV et sous-estimer leurs symptômes (FDA, 2009; Roydhouse et al., 2019). La seconde problématique soulevée par les essais en ouvert est le risque de biais des résultats lié au remplissage des questionnaires (Roydhouse et al., 2019). En raison de la déception possible d'être randomisé dans le bras contrôle, les patients pourraient être moins susceptibles de répondre aux questionnaires que ceux randomisés dans le bras expérimental. La FDA a cherché à comparer les taux de compliance des données de PROs entre les essais cliniques randomisés en ouvert et en double aveugle (Roydhouse, 2019). Par le biais de leur étude, ils ont mis en évidence que les taux de compliance étaient élevés quel que soit le type d'étude. Cependant, dans le peu d'essais où une légère différence était observée, celle-ci favorisait en particulier le bras expérimental et était plus courante dans les essais en ouvert que dans les essais en double aveugle.

Enfin, l'évaluation de la QdV à l'inclusion est un point clé dans l'analyse longitudinale. Cette dernière devrait systématiquement être évaluée avant toute randomisation ou du moins avant l'initiation du traitement, comme le recommande le groupe EORTC (Kiebert et al., 1998). En effet, lorsque l'évaluation à l'inclusion arrive après la randomisation dans des essais en

ouvert, les patients connaissent le traitement attribué ce qui peut biaiser les résultats. Or, l'évaluation à l'inclusion permet de pouvoir fournir un point de référence essentiel pour identifier les changements dans le temps.

Ces dernières années, la FDA a accordé l'inclusion d'informations sur les PROs dans le résumé des caractéristiques des produits à très peu de molécules (Gnanasakthy et al., 2019). La FDA étudie activement les biais pouvant être observés dans le cadre des essais en ouvert. La prise en compte des biais pour les résultats de PROs dans les essais en ouvert est importante. Le biais lié à la connaissance de l'attribution d'un traitement n'est pas spécifique aux données de QdV, elle peut également affecter d'autres résultats d'essais cliniques, comme par exemple la survie sans progression (Roydhouse et al., 2019). En effet, dans les essais cliniques randomisés en ouvert, l'évaluation de la progression par l'investigateur ou le radiologue peut être biaisée par une préférence de traitement (Sridhara et al., 2013).

II. Objectifs

Les PROs, incluant la QdV, sont reconnus comme des critères de jugement majeurs dans les essais cliniques en cancérologie. Ces critères permettent d'intégrer la voix des patients dans l'évaluation des risques et des bénéfices cliniques des nouvelles stratégies de traitement.

Bien que les PROs soient désormais fréquemment collectés dans les essais cliniques en cancérologie, l'analyse et l'interprétation de ces données restent complexes et non standardisées, ce qui compromet la comparaison des résultats entre les essais cliniques. L'occurrence des données manquantes fait partie des enjeux majeurs de l'analyse longitudinale des données de PROs. Les données manquantes peuvent être informatives de l'état de santé et du niveau de QdV du patient. Elles peuvent affecter l'analyse longitudinale en réduisant la taille de l'échantillon et la puissance statistique. Ces dernières peuvent également entraîner un biais dans les analyses et les résultats si elles ne sont pas correctement prises en compte. L'apparition d'un effet « *response shift* » s'ajoute également aux enjeux majeurs de l'analyse longitudinale des données de PROs. Les patients peuvent évaluer différemment leur QdV au cours du temps en raison de différents facteurs. Ainsi, ignorer l'occurrence d'un effet « *response shift* » pourrait impacter et biaiser les résultats.

Par ailleurs, parmi les nombreuses méthodes proposées pour l'analyse longitudinale des données de PROs, figure l'approche du temps jusqu'à détérioration. Cette dernière méthode a été proposée récemment et est régulièrement utilisée comme modalité d'analyse longitudinale dans les essais cliniques de phase III. L'approche du temps jusqu'à détérioration nécessite une définition claire de ce qui est considéré comme une « détérioration ». Elle dépend de nombreux critères tels que la localisation cancéreuse et la situation thérapeutique, le score de référence, la DMCI perçue par le patient, ainsi que les règles de censure. Cette méthode d'analyse longitudinale propose des résultats facilement compréhensibles et interprétables pour les cliniciens, déjà familiarisés avec les méthodes de type « temps jusqu'à événement ». Cependant, cette méthode est utilisée de manière non standardisée, due à la variabilité des définitions du temps jusqu'à détérioration existantes et des événements multiples potentiellement inclus dans la définition de l'événement, donnant lieu à des critères dits composites.

Dans cette optique, les objectifs des différents travaux ont été centrés sur la méthode du temps jusqu'à détérioration en tant que modalité d'analyse longitudinale des données de PROs. Dans

un premier temps, une revue systématique de la littérature dans les essais cliniques randomisés de phase III a été réalisée afin de faire l'état des définitions du temps jusqu'à détérioration utilisées. Dans un second temps, l'ensemble des définitions recommandées du temps jusqu'à détérioration a été implémenté sous le logiciel SAS afin d'optimiser et d'harmoniser l'utilisation de cette méthode. Ces différentes implémentations permettront une diffusion et une utilisation plus large de cette méthode afin de progresser vers l'objectif de standardisation de l'analyse longitudinale des données de PROs dans les essais cliniques en cancérologie. Puis, la méthode du temps jusqu'à détérioration a été étudiée dans une cohorte de patientes atteintes d'un cancer du sein en situation adjuvante. Cette étude a conduit à s'intéresser à la première détérioration du patient, à la gestion de l'absence de randomisation à l'inclusion ainsi qu'à l'impact potentiel d'un effet « *response shift* ». Enfin, la méthode du temps jusqu'à détérioration a également été explorée dans un essai de phase II randomisé chez des patients atteints d'un cancer du pancréas métastatique. Dans cet essai, l'impact de l'occurrence des données manquantes à l'inclusion a été traité.

III. Résultats

1. Revue de la littérature sur la méthode du temps jusqu'à détérioration

<u>Article :</u> Time to deterioration in cancer randomized clinical trials for patient-reported outcomes data: a systematic review.

Article publié dans *Quality of Life Research*

Résumé

Objectif

L'approche du temps jusqu'à détérioration a été proposée comme modalité d'analyse longitudinale des PROs dans les essais cliniques randomisés en cancérologie. L'objectif de cette étude était de réaliser une revue systématique de la littérature sur la manière dont l'approche du temps jusqu'à détérioration a été définie et rapportée dans les essais cliniques randomisés de phase III depuis 2014 pour analyser les données de PROs. En particulier, le niveau de clarté de la définition du temps jusqu'à détérioration utilisée a été examiné dans les études. La concordance a également été étudiée entre les résultats des données de PROs analysés selon l'approche du temps jusqu'à détérioration et les résultats du (des) critère(s) de jugement principal (principaux) de l'étude.

Méthodes

Une recherche systématique de la littérature a été menée sur les bases de données PubMed/MEDLINE, la bibliothèque Cochrane et par une recherche manuelle, afin d'identifier les études publiées entre Janvier 2014 et Juin 2018. Tous les essais cliniques randomisés de phase III en cancérologie, incluant les PROs comme critère de jugement et utilisant l'approche du temps jusqu'à détérioration, ont été considérés. Nous avons collecté des informations relatives à l'étude, à l'évaluation des PROs et à l'approche du temps jusqu'à détérioration. Le niveau de clarté de la définition du temps jusqu'à détérioration utilisée a été évalué par le biais de cinq critères clés qui devraient être reportés dans les publications pour comprendre la définition employée. Les cinq critères clés sont : (1) la définition de la

détérioration, (2) le score de référence utilisé pour qualifier la détérioration, (3) la DMCI, (4) si une définition composite a été utilisée et (5) la manière dont les données manquantes à l'inclusion ont été traitées. Chaque critère a été codé comme « défini » (2 points), « pas clair » (1 point) ou « non défini » (0 point). Ainsi, un score allant de 0 à 10 a été créé, plus le score était élevé plus la définition était précise. Les résultats obtenus pour le(s) critère(s) de jugement principal (principaux) ont été recueillis et classés comme étant « en faveur du bras expérimental », « en faveur du bras contrôle », « pas clair » (i.e. impossible d'identifier la direction dans laquelle les résultats allaient) ou « aucune différence » (i.e. absence de différence statistiquement significative entre les deux bras de traitement), selon la signification statistique rapportée dans les résultats. La même classification a été effectuée pour les résultats des données de PROs. Les résultats ont été classés comme étant « en faveur du bras expérimental » ou « en faveur du bras contrôle » pour les résultats des données de PROs, si plus de la moitié des dimensions ciblées/attendues statistiquement significatives étaient respectivement en faveur du bras expérimental ou du bras contrôle.

Résultats

Parmi les 1549 articles examinés, 39 études ont été identifiées comme pertinentes selon des critères prédéfinis. Parmi ces études, 36 études (92,3%) portaient sur des cancers en situation avancée et/ou métastatique. Sept études (17,9%) ont considéré un co-critère de jugement principal, correspondant à la survie globale et la survie sans progression. Les PROs ont été considérés comme un critère de jugement secondaire dans 32 études (82,1%), exploratoire dans 6 études (15,3%) et 1 étude (2,6%) n'a pas clairement défini le statut du critère de jugement des PROs. Plusieurs définitions de détérioration ont été utilisées dans ces essais : 10 études (25,6%) ont défini la détérioration comme « définitive », correspondant à une détérioration maintenue jusqu'à la dernière évaluation de PROs disponible pour chaque patient. Elle était « confirmée » dans 7 études (17,9%), correspondant à une détérioration maintenue pendant une période de temps définie, qui pourrait être égale ou inférieure à l'intervalle de temps défini dans la définition précédente. La détérioration a été explicitement déclarée comme une première détérioration dans 16 études (41,0%) et sous-entendue dans 8 études (20,5%). La définition de la détérioration n'a pas été clairement définie pour 1 étude (2,6%). Le score à l'inclusion a été clairement déclaré comme le score de référence pour qualifier la détérioration dans la plupart des études (n = 31, 79,5%). La médiane du niveau de

clarté était de 7 (étendue 5-9) lorsque les résultats des données de PROs étaient rapportés dans la publication principale (n = 7, 17,9%) et 9 (étendue 4-10) pour les publications secondaires consacrées aux résultats des données de PROs.

Conclusion

Cette revue souligne le manque de standardisation de l'approche du temps jusqu'à détérioration pour l'analyse des données de PROs dans les essais cliniques randomisés. Enfin, la définition de la « détérioration » devrait être adaptée plus systématiquement, à la situation thérapeutique.

Quality of Life Research
https://doi.org/10.1007/s11136-019-02367-7

REVIEW

Time to deterioration in cancer randomized clinical trials for patient-reported outcomes data: a systematic review

E. Charton[1,2] · B. Cuer[3,4] · F. Cottone[5] · F. Efficace[5] · C. Touraine[3] · Z. Hamidou[6,7] · F. Fiteni[3,8] · F. Bonnetain[1,2,7] · M.-C. Woronoff-Lemsi[2,9] · C. Bascoul-Mollevi[3,4,7] · A. Anota[1,2,7]

Accepted: 18 November 2019
© Springer Nature Switzerland AG 2019

Abstract

Purpose The time to deterioration (TTD) approach has been proposed as a modality of longitudinal analysis of patient-reported outcomes (PROs) in cancer randomized clinical trials (RCTs). The objective of this study was to perform a systematic review of how the TTD approach has been used in phase III RCTs to analyze longitudinal PRO data.

Methods A systematic literature search was conducted in PubMed/MEDLINE, the Cochrane Library and through manual search to identify studies published between January 2014 and June 2018. All phase III cancer RCTs including a PRO endpoint using the TTD approach were considered. We collected general information about the study, PRO assessment and the TTD approach, such as the event definition, the choice of reference score and whether the deterioration was definitive or not.

Results A total of 1549 articles were screened, and 39 studies were finally identified as relevant according to predefined criteria. Among these 39 studies, 36 (92.3%) were in advanced and/or metastatic cancer. Several different deterioration definitions were used in RCTs, 10 studies (25.6%) defined the deterioration as "definitive", corresponding to a deterioration maintained over time until the last PRO assessment available for each patient. The baseline score was explicitly stated as the reference score to qualify the deterioration for most studies ($n=31$, 79.5%).

Conclusion This review highlights the lack of standardization of the TTD approach for the analysis of PRO data in RCTs. Special attention should be paid to the definition of "deterioration", and this should be based on the specific cancer setting.

Keywords Patient-reported outcomes · Randomized clinical trials · Time to deterioration · Systematic review

Introduction

Patient-reported outcomes (PROs), including health-related quality of life (HRQoL), are key endpoints in oncology clinical trials to assess the clinical benefit of new treatment strategies for the patients [1, 2]. However, the analysis of

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11136-019-02367-7) contains supplementary material, which is available to authorized users.

✉ E. Charton
echarton@chu-besancon.fr

1 Methodology and Quality of Life Unit in Oncology (INSERM UMR 1098), University Hospital of Besançon, 3 Boulevard Fleming, 25030 Besançon, France

2 University Bourgogne Franche-Comté, INSERM, EFS BFC, UMR1098, Interactions Hôte-Greffon-Tumeur/Ingénierie Cellulaire et Génique, Besançon, France

3 Montpellier Cancer Institute (ICM), University of Montpellier, Montpellier, France

4 Institute of Cancer Research of Montpellier (IRCM), ICM, INSERM, Montpellier, France

5 Data Center and Health Outcomes Research Unit, Italian Group for Adult Hematologic Diseases (GIMEMA), Rome, Italy

6 EA3279 Self-perceived Health Assessment Research Unit, Aix-Marseille University, Marseille, France

7 French National Platform Quality of Life and Cancer, Besançon, France

8 Department of Medical Oncology, University Hospital of Nîmes, Nîmes, France

9 Department of Clinical Research and Innovation, University Hospital of Besançon, Besançon, France

Published online: 27 November 2019

PRO data remains challenging [3], as no standardization of statistical methods has yet been proposed, thus rendering comparison of PRO results between clinical trials difficult. Several methods have been proposed for the longitudinal analysis of PRO data including the linear mixed model [4, 5] and the time to deterioration (TTD) approach [6–8]. This latter method has recently been proposed and is regularly used as a modality of longitudinal analysis in phase III clinical trials [3]. The TTD approach requires a clear definition of what is considered to be a "deterioration", and this will depend on the cancer setting, the reference score and the minimal important difference (MID) used to qualify the deterioration, as well as on censoring rules. The deterioration could be considered as "reversible", "definitive" (i.e., a deterioration maintained over time up until the last PRO assessment available for each patient), or "confirmed" (i.e., a deterioration sustained for a defined time period, which might be equal to or less than the time lapse defined in the preceding definition). Composite events may also be considered, for example, including death in the event definition. Since the variability in the TTD definitions may have an impact on the comparison of results between trials, some initial recommendations have previously been proposed in a paper published online in 2013, and they emphasized the need to adapt the definition based on the therapeutic setting [6]. For example, in the adjuvant setting, it has been proposed to use the time to first clinically significant deterioration as compared to the baseline score ("reversible" deterioration) [8]. In contrast, in advanced or metastatic disease, the time until definitive deterioration (whether or not death is included as an event) seemed to be more appropriate. The recommended definition of the time until definitive deterioration is the time to the first clinically significant deterioration compared to the baseline score, and with no further significant improvement compared to the baseline score ("definitive" deterioration) [7].

Despite these initial recommendations, however, various other definitions have been used in randomized clinical trials (RCTs) [9, 10], thereby limiting comparisons of results across RCTs, as with other traditional time-to-event endpoints, such as progression-free survival [11].

The objective of this study was to perform a systematic review of how the TTD approach has been used in phase III RCTs to analyze longitudinal PRO data. The primary objective was to examine the level of clarity of the TTD definition used. The secondary objective was to assess the concordance between PRO data using the TTD approach, and the primary endpoint(s), in the studies included in the review.

Methods

Search methods for identification of studies

A systematic literature search was conducted in PubMed/MEDLINE, the Cochrane Library and through manual searching since 2014. The search strategies combined different terms to represent PROs, quality of life and the TTD approach. A filter for RCTs was performed on PubMed/MEDLINE, using the Cochrane highly sensitive search strategy-sensitivity maximizing version (2008 revision), as described in the Cochrane Handbook for Systematic Reviews of Interventions [12]. We list the full search strategies for PubMed/MEDLINE and the Cochrane Library in Supplementary Table A.

Criteria for considering studies

All phase III RCTs in oncology published between January 2014 and June 2018 were eligible if they included a PRO endpoint and used the TTD approach to analyze PRO data. Among the articles identified, only original articles written in English were considered. We chose 2014 as the starting publication date for studies to be included in this review, as it was the year after the online publication (in 2013) of the initial article proposing recommendations for the TTD approach [6].

Selection of studies

Two authors (E.C., B.C.) independently assessed the eligibility of all articles identified by the searches. Articles that did not meet the inclusion criteria after screening the titles and abstracts or the full text were excluded. The reasons for exclusion of the ineligible studies were recorded. Disagreements were resolved through discussion between the two reviewers. In case of unresolved disagreements, a third author (A.A.) was consulted. Duplicates studies were identified and excluded, and only the original article(s) reporting PROs data by treatment arm were selected.

Data extraction

Two authors (E.C., A.A.) independently extracted information defined in a data collection form for each selected study. All extracted data were cross-checked, and discrepancies were resolved after discussion between the two reviewers.

The data extraction form recorded the following:

- general information about the study and PRO assessment, i.e., year of publication, study location (i.e., the trial was considered as an international study if it recruited patients in more than one country), publication status (i.e., secondary publication dedicated to PRO results versus PRO results presented in the main publication), cancer site, disease stage, primary endpoint, PRO endpoint status, reporting of the baseline PRO sample size, PRO population considered for the analyses, questionnaire(s) used and timing of PRO assessments;
- items specific to the TTD approach: the definition of deterioration used, the reference score used to qualify the deterioration, the MID considered, whether composite events were considered and if so, the events included in the composite, missing data and sensitivity analysis;
- the reporting of the results, e.g., description of the targeted/expected baseline PRO scores, results presentation and reported events. If the study did not explicitly state that it focused only on specific dimensions of the questionnaire(s) used ("targeted dimensions"), then we considered all dimensions of the questionnaire(s) used, which we called the "expected dimensions".

For the primary objective of this review, the clarity of the TTD definition used was assessed separately for main publications and for secondary publications dedicated to PRO results, according to 5 key criteria, namely: (1) the definition of deterioration (e.g., first deterioration, deterioration sustained for a defined time period, deterioration maintained over time); (2) the reference score used to qualify the deterioration; (3) the MID; (4) whether a composite definition was used (and if so, the events included in the composite); and (5) the manner in which missing baseline data were handled. Each criterion was coded as "defined" (2 points), "unclear" (1 point) or "not defined" (0 point). A score ranging from 0 to 10 was thus created, with a higher score indicating greater clarity of the TTD definition used.

The secondary objective regarding the concordance between PRO data using TTD approach and the primary endpoint(s) of the study was also investigated. Results obtained for the primary endpoint(s) were collected and classified as being "in favor of the experimental arm", "in favor of the control arm", "unclear" (i.e., impossible to identify which direction the results tended) or "no difference" (i.e., clear statement of an absence of statistically significant difference), according to the statistical significance reported in the results. The same classification was performed for the PRO endpoint using the TTD results. Results were classified as being "in favor of the experimental arm" or "in favor of the control arm" for the PRO endpoint, if more than half of the statistically significant targeted/expected dimensions were in favor of the experimental or control arm, respectively.

Data analysis

We performed descriptive analyses of the articles identified. Qualitative variables are described as absolute and relative frequencies, while quantitative variables are described as median and range.

Analyses were conducted using SAS version 9.3 (SAS Institute Inc., Cary, NC, USA).

Results

The systematic literature review search yielded 1549 studies published between January 2014 and June 2018, which were screened for eligibility. After duplicates were removed, 1229 records were screened and a total of 39 studies (2.5%) were finally identified as relevant according to the predefined inclusion criteria (Fig. 1 and Supplementary Table B) [9, 10, 13–49].

General information about selected studies

Out of the 39 selected studies, the majority were international ($n=32$, 82.1%) and most of the time, a secondary publication dedicated to the PRO results was provided ($n=32$, 82.1%) (Table 1). The two main cancer sites were lung ($n=12$, 30.8%) and gastrointestinal ($n=11$, 28.2%) cancer. Thirty-six studies (92.3%) were on advanced and/or metastatic cancer and 3 studies (7.7%) on localized cancer only. Seven studies (17.95%) investigated a co-primary endpoint, considering overall and progression-free survival. PRO was considered as a secondary endpoint in 32 studies (82.1%), an exploratory endpoint in 6 studies (15.3%), and 1 study (2.6%) did not clearly define the status of the PRO endpoint. The population used for the longitudinal analysis of PRO was defined as intention-to-treat in 12 studies (30.8%), as a modified intention-to-treat in 19 studies (48.7%) was not clearly defined in 3 studies (7.7%) and was not reported in 5 studies (12.8%). The definition of the modified intention-to-treat PRO population varied by article. Regarding how missing baseline scores were considered, while an intention-to-treat population was declared for the analysis, only 2 studies (16.7%) considered observations with missing baseline data, 1 study (8.3%) had no missing baseline scores, 5 studies (41.7%) did not clearly define how missing baseline data were handled, and 4 studies (33.3%) did not report this information. Among all studies, the European Organisation for

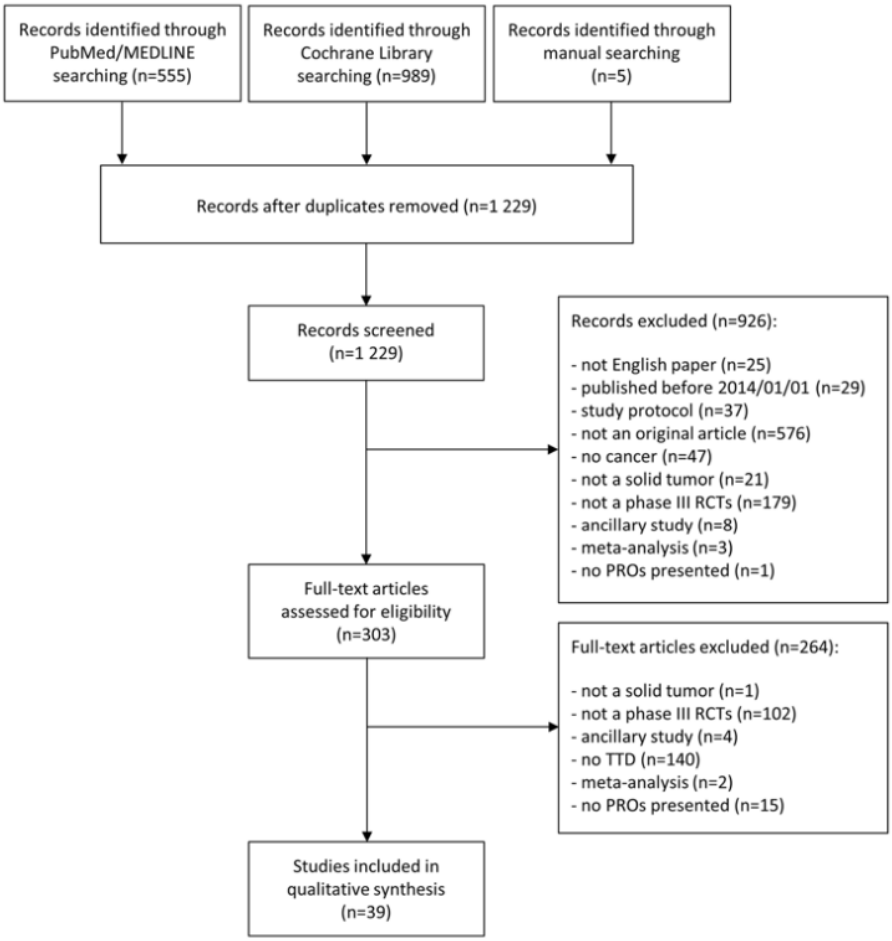


Fig. 1 Flow of information through the different phases of the systematic review. *RCTs* randomized clinical trials, *PROs* patient-reported outcomes, *TTD* time to deterioration

Research and Treatment of Cancer (EORTC) questionnaires were the most frequently used instruments ($n=26$, 66.7%).

Methodology of the time to deterioration approach

The deterioration was defined as definitive in 10 studies (25.6%), corresponding to a deterioration maintained over time until the last PROs assessment available for each patient, and confirmed in 7 studies (17.9%), which corresponds to a deterioration sustained for a defined time period that could be equal to or less than the time lapse defined in the preceding definition (Table 2). The deterioration was explicitly stated as a first deterioration in 16 studies (41.0%) and was not explicitly stated as a first deterioration but implied in 8 studies (20.5%). The deterioration definition was not clearly defined in 1 study (2.6%). Two studies (5.1%) applied several definitions of deterioration in their main analysis [20, 24]. The baseline score was explicitly stated as the reference score to qualify the deterioration in most studies ($n=31$, 79.5%) and was either not defined or unclear for the remaining 8 studies (20.5%). The MID was defined for all targeted/expected dimensions in 35 studies (89.7%), for some dimensions in 3 studies (7.7%) and was not defined in one study (2.6%). This MID was generally the same for each dimension ($n=29$, 76.3%), and only 8 studies (21.1%) used a different MID according to the dimension. Composite definitions of PRO deterioration were considered in 16 studies (41.0%), including deterioration on any of three PRO scales ($n=1$, 6.3%), an increase in analgesic use ($n=1$, 6.3%), death ($n=14$, 87.5%) and disease progression ($n=5$, 31.3%) in the event definition. Five studies (31.3%) simultaneously considered death and disease progression as composite definitions. The majority of studies did not report how intermittent ($n=37$, 94.9%) or monotone ($n=24$, 61.5%) missing data were handled. Regarding studies in advanced and/or metastatic cancer ($n=36$, 92.3%), only 12 studies (33.3%) considered death in the deterioration definition in their main analysis. Among the 24 studies (66.7%) that

Table 1 General characteristics of the studies included ($n = 39$)

	n	%
Year of publication		
2014	4	10.3
2015	10	25.6
2016	9	23.1
2017	8	20.5
2018	8	20.5
International study[a]		
Yes	32	82.1
No	7	17.9
Publication status		
PRO results presented in the main publication	7	17.9
Secondary publication dedicated to PRO results	32	82.1
Cancer site		
Brain cancer	2	5.1
Breast cancer	4	10.3
Gastrointestinal cancer	11	28.2
Head and neck cancer	2	5.1
Lung cancer	12	30.8
Melanoma	2	5.1
Urological cancer	7	17.9
Disease stage		
Advanced/metastatic	36	92.3
Local	3	7.7
Primary endpoint(s)		
2-Year disease-free survival	1	2.6
Overall survival	20	51.3
Progression-free survival	24	61.5
Time-to-treatment failure	1	2.6
PRO endpoint		
Secondary	32	82.1
Exploratory	6	15.3
Not reported	1	2.6
Baseline PRO sample size		
Reported	31	79.5
Not reported/unclear	8	20.5
PRO population for the analyses		
Intention-to-treat	12	30.8
Modified intention-to-treat	19	48.7
Unclear	3	7.7
Not reported	5	12.8
PRO questionnaires		
Brief Fatigue Inventory	1	2.6
Brief Pain Inventory-Short Form	4	10.3
EQ-5D questionnaires	15	38.5
EORTC questionnaires	26	66.7
FACT questionnaires	8	20.5
Lung Cancer Symptom Scale	3	7.7
Visual Analogue Scale	1	2.6
Number of PRO assessments		
=4	1	2.6
>4	31	79.5
Not reported	2	5.1
Unclear	5	12.8
Length of PRO assessment		
Up to 6 months after inclusion in the study	1	2.6
Up to 1 year	9	23.1
More than 1 year	20	51.3
Unknown	3	7.7
Unclear	6	15.3

Total percentages may exceed 100 since several responses were possible in each study

PRO patient-reported outcome, *EQ-5D* EuroQoL-5 Dimensions, *EORTC* European Organisation for Research and Treatment of Cancer, *FACT* Functional Assessment of Cancer Therapy

[a]The trial was considered as an international study if it recruited patients in more than one country

did not consider death in the main analysis, only 1 study (4.2%) integrated it in a sensitivity analysis.

Examination of the clarity of the time to deterioration definition used

The median level of clarity based on the 5 key criteria was 7 (range 5–9) when PROs were reported in the main publication ($n = 7$, 17.9%) and 9 (range 4–10) for the secondary publications dedicated to PRO results ($n = 32$, 82.1%). Regarding papers reporting PRO data in the main publication, 4 studies (57.1%) defined the deterioration definition, 3 studies (42.9%) defined the reference score, all studies defined both the MID and composite events, and none defined how missing baseline scores were handled. Regarding the 32 secondary publications, 26 studies (81.2%) defined the deterioration, 28 studies (87.5%) defined the reference score and the MID, all studies defined composite events and 17 studies (53.1%) described their method of handling missing baseline scores (Table 3).

Reporting of the results

The targeted/expected baseline PRO scores were reported for all scores in 23 studies (59.0%), for some scores in 3 studies (7.7%) and were not reported in 13 studies (33.3%). Most of the studies presented the results for all targeted/expected dimensions ($n = 33$, 84.7%), and 6 studies (15.3%) presented only the results for some dimensions. The number of events reported for each targeted/expected dimension was not mainly reported ($n = 24$, 61.5%). The

Table 2 Definition of the time to deterioration approach considered in the studies included ($n=39$)

	n	%
Definition of deterioration		
First deterioration explicitly stated	16	41.0
First deterioration not explicitly stated but implied	8	20.5
Definitive deterioration (maintained over time)	10	25.6
Confirmed deterioration (sustained for a defined time period)	7	17.9
Unclear	1	2.6
Reference score to qualify the deterioration		
Baseline	31	79.5
Not defined	5	12.8
Unclear	3	7.7
Minimal important difference defined		
Yes, for all scores	35	89.7
Yes, for some scores	3	7.7
No	1	2.6
Consideration of the minimal important difference ($n=38$)		
Same value for each scale	29	76.3
Different according to the scale	8	21.1
Different according to the patient	2	5.3
Unclear	1	2.6
Composite definition		
Yes	16	41.0
No	23	59.0
Composite events ($n=16$)		
Deterioration in any of three PRO scales	1	6.3
Death	14	87.5
Disease progression	5	31.3
Increase in analgesic use	1	6.3
Consideration of baseline missing scores		
Censor	2	5.1
Not reported	9	23.1
Unclear	13	33.3
Not applicable[a]	15	38.5
Intermittent missing data		
Deterioration occurred at the time of the missing value	2	5.1
Not reported	37	94.9
Monotone missing data		
Deterioration occurred at the time of the missing value	1	2.6
Death ignored	1	2.6
Death considered as an event	14	35.9
Not reported	24	61.5
Sensitivity analyses		
Yes (stated in the article)	10	25.6
No	26	66.7
Unclear (not clearly stated, but considered as such)	3	7.7
Sensitivity analyses performed ($n=13$)		
Minimal important difference	6	46.2
Death considered as an event	3	23.1
Death excluded as an event	2	15.4
Disease progression considered as a competing risk	1	7.7
Disease progression excluded as an event	2	15.4
Only data up to disease progression	1	7.7
First deterioration definition applied	1	7.7
Missing data	1	7.7
Unclear	1	7.7

Total percentages may exceed 100 since several responses were possible in each study

PRO patient-reported outcome

[a]Not applicable, if modified intention-to-treat population or no missing baseline scores

same results were observed in case of composite definitions ($n=12$, 75.0%) (Table 4).

Concordance between PRO data using the time to deterioration approach and the primary endpoint(s)

Among all studies involved, the primary endpoint(s) was (were) either in favor of the experimental arm, or did not differ between treatment arms. The results obtained for the primary endpoint(s) and the PRO endpoint were concordant in 27 studies (69.2%). Among these, results were in favor of the experimental arm in 23 studies (85.2%) and found no difference between treatment arms in 4 studies (14.8%) (Table 5).

Discussion

This review aimed to study how the TTD approach has been used in phase III RCTs to analyze longitudinal PRO data. We identified 39 studies published between January 2014 and June 2018 that used the TTD approach. A number of studies used a composite definition of deterioration (41.0%), mostly including death (with or without disease progression) in the composite event. The choice of the events associated in the TTD definition should be made carefully and must be justified. Indeed, including disease progression in the TTD definition could be controversial, since disease progression is a tumor-centered endpoint and not a patient-centered endpoint [50]. In this review, 5 studies included disease progression in the TTD definition. In a composite endpoint, each component should be of similar importance to the patient and thus be clinically relevant [51], and homogeneous treatment effects on each component should be investigated [3, 52]. However, HRQoL is generally assessed until disease

Table 3 The 5 key criteria used to assess the clarity of the time to deterioration definition used in the studies included (n=39)

	Main publication (n=7)			Secondary publication dedicated to PRO results (n=32)		
	Defined, n (%)	Unclear, n (%)	Not defined, n (%)	Defined, n (%)	Unclear, n (%)	Not defined, n (%)
Definition of the deterioration	4 (57.1)	3 (42.9)	0 (0.0)	26 (81.2)	6 (18.8)	0 (0.0)
Reference score to qualify the deterioration	3 (42.9)	1 (14.2)	3 (42.9)	28 (87.5)	2 (6.3)	2 (6.3)
Minimal important difference	7 (100.0)	0 (0.0)	0 (0.0)	28 (87.5)	3 (9.4)	1 (3.1)
Composite definitions	7 (100.0)	0 (0.0)	0 (0.0)	32 (100.0)	0 (0.0)	0 (0.0)
Method of handling missing baseline scores	0 (0.0)	3 (42.9)	4 (57.1)	17 (53.1)	10 (31.3)	5 (15.6)

PRO patient-reported outcome

Table 4 Reporting of the patient-reported outcome (PRO) results of the studies included (n=39)

	n	%
Description of the targeted/expected baseline PRO scores		
Yes, for all targeted/expected scores	23	59.0
Yes, for some scores	3	7.7
No	13	33.3
Results presented for targeted/expected dimensions		
Yes, for all targeted/expected scores	33	84.7
Yes, for some scores	6	15.3
Reporting of the deterioration results		
Deterioration curve by treatment arm	27	69.2
Forest plots	10	25.6
Hazard ratio of the treatment effect with confidence interval	34	87.2
Hazard ratio of the treatment effect without confidence interval	1	2.6
Median with confidence interval	20	51.3
Median without confidence interval	12	30.8
Plot of differences in median by treatment arm	1	2.6
Number of events reported for each targeted/expected dimension		
Yes, for all targeted/expected scores	14	35.9
Yes, for some scores	1	2.6
No	24	61.5
Number of events reported for each component in case of composite definitions (n=16)		
Yes	3	18.7
No	12	75.0
Unclear	1	6.3

Total percentages may exceed 100 since several responses were possible in each study

Table 5 Concordance of the results between patient-reported outcome (PRO) data using the time to deterioration approach and the primary endpoint(s)

PRO endpoint	Primary endpoint(s)		
	In favor of the experimental arm, n (%)	No difference, n (%)	Total, n (%)
In favor of the experimental arm	23 (79.3)[a]	5 (50.0)	28 (71.8)
In favor of the control arm	0 (0.0)	1 (10.0)	1 (2.6)
No difference	5 (17.2)	4 (40.0)[a]	9 (23.0)
Unclear	1 (3.5)	0 (0.0)	1 (2.6)
Total	29 (74.4)	10 (25.6)	39 (100.0)

[a]Concordance between PROs and primary endpoint(s)

progression in RCTs. In such a case, ignoring disease progression would result in informative censoring.

In this review, different definitions of deterioration were reported, and the definitions differed even between two similar clinical trials. For example, two RCTs were performed in metastatic castration-resistant prostate cancer, comparing enzalutamide to placebo [24, 34]. Both trials assessed PRO using the Functional Assessment of Cancer Therapy-Prostate (FACT-P) and Brief Pain Inventory-Short Form questionnaires and analyzed the deterioration of the FACT-P total score as well as pain progression. In the AFFIRM trial [24], the FACT-P total score deterioration was defined as a first deterioration of at least 10 points compared to the baseline score, or death from any cause. In the PREVAIL trial [34], the FACT-P total score deterioration was defined as a first deterioration of at least 10 points compared to the baseline score. Thus, death was not considered as an event in the TTD definition in this latter study. Consequently, results cannot be directly compared between these two trials. Since the TTD approach seems to be used regularly to analyze longitudinal PRO data, it is mandatory to propose a standardization of the deterioration definition to be used, that is adapted to the cancer site and setting (localized or advanced/metastatic) and easy to follow in practice. Some initial recommendations were published online in 2013 [6], proposing a standardized definition by cancer setting (adjuvant or advanced/metastatic setting). In metastatic cancer, a definitive deterioration, whether or not it includes death as an event, was the recommended definition. However, the AFFIRM and PREVAIL trials both failed to follow these recommendations. While the majority of the studies in this review were performed in the advanced/metastatic cancer (92.3%), the deterioration considered was mostly a first deterioration, whether this was explicitly stated (44.4%) or not (22.2%). In contrast, the deterioration was defined as definitive or confirmed in 19.4% and 19.4% of studies, respectively. In general, very few papers have considered the 2013 recommendations [6]. Due to the low number of papers published per year, it was not possible to investigate whether any improvement in adherence to the recommendations occurred over time. This failure to take these recommendations into account may have several explanations. Firstly, communication about this methodological work may be insufficient. Second, researchers may consider these recommendations inappropriate for their setting. Third, the statistical analysis plans of the trials, including the definition of TTD, may have been written prior to the online publication of the recommendations in 2013. Finally, this could be due to the complexity of performing the analyses. Indeed, investigating a definitive deterioration of HRQoL requires checking that the deterioration is maintained at all subsequent measurement times. In order to facilitate the application of the TTD approach and to allow some standardization, this approach has been implemented in R software [53]. Moreover, the development of international guidelines is ongoing with the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data (SISAQOL) project and will include the TTD approach [54].

At this time, no recommendation has been made on how death should be taken into account. In this review, 87.5% of the studies considered death as an event. To ignore death in advanced cancer could represent informative censoring. However, if death occurs a long time after the last available HRQoL assessment, it could overestimate the TTD and may be a better reflection of overall survival if most of the observed events are due to death. One could thus include death as an event if it occurred in a reasonable length of time after the last HRQoL assessment [55]. An alternative would be to consider death as a competing risk with the event of interest, i.e., the deterioration of HRQoL [45].

The use of the TTD approach requires specifying important details. Hence, the primary objective of this review was to assess the clarity of the TTD definition according to 5 key criteria. Not surprisingly, the clarity of the TTD definition was slightly better in the secondary papers than when PRO results were reported in the main publication. Overall, the clarity score was rather high for all studies. However, since only seven main publications were identified, no statistical comparison with the secondary publications dedicated to PRO results was performed. The manner of handling missing baseline scores was the least well-defined issue. A majority of studies did not clearly report how missing data at baseline were considered. Only two studies considered missing baseline scores as a censor [20, 31]. In the TTD approach, the baseline score is a key point. In case of missing baseline scores, it is necessary to know how its data are considered in order to apply the TTD approach. Most of the time, in this review, there was no mention of how intermittent or monotone missing data were handled over time. Only 2 studies assumed that the deterioration occurred at the time of the missing value [24, 44], which could make it possible to take account of informative missing data.

The secondary objective of this review was to assess the concordance between PRO data using the TTD approach and the primary endpoint(s) of the study. This objective highlighted the consistency and complementarity of the results in 27 studies (69.2%).

The TTD is strongly influenced by the measurement times. Although the measurement times are generally similar between treatment arms, the frequency and interval between two consecutive assessments can vary between RCTs. Most of the clinical trials in this review assessed PROs at every treatment cycle. However, some of them assessed PROs every 3 months [36] or every 3 and then every 6 months

[45]. Furthermore, the assessment time may depend on the disease course and the expected change in HRQoL. The median TTD could be overestimated if the measurement times are too far apart. This is a common problem for time-to-event endpoints [56], and interval censoring could be a feasible solution to be investigated. The TTD approach should be applied only with a reasonable number of measurement times [3]. The EORTC recommends at least three HRQoL assessments in clinical trials [57] in order to assess the impact of treatment on HRQoL over time. To apply the TTD, three measurement times seem to be the strict minimum. We chose to report here the number of studies with at least four measurement times. In this review, most studies fulfilled this criterion.

In future research, to consider the PRO as a primary or co-primary endpoint using the TTD approach, certain assumptions have to be made [58]. These assumptions should be based on TTD data from previous studies. Furthermore, assuming that the measurement times can impact this approach, the timing of the PRO assessments should be similar to that used in the prior study on which the TTD assumptions are based. Indeed, more intensive assessment of PRO in one arm compared to the other could introduce bias in the TTD. In particular, if there are more missing data in one arm than in the other, then the TTD could be overestimated in the arm with more missing data. If it is not possible to make assumptions about the TTD, then TTD should not be used as the primary endpoint for PROs. Alternative approaches should be preferred, such as the mean difference in PRO scores measured at two different timepoints, or the difference between treatment arms, according to the MID. In any case, the PRO questionnaire should be administered at the earliest time(s) when MIDs are expected to occur.

This literature review has several limitations. First, it was limited to 4 years, from 2014 to 2018. This was due to the fact that the initial paper providing recommendations of the deterioration definition was published in late 2013. The objective was thus to observe whether the papers published since 2014 followed these initial recommendations. However, this review covers a short period and it would be interesting to reiterate this work in a few years from now. Indeed, this methodology is almost a novel approach to analyzing PRO data.

Conclusion

This systematic review highlights the heterogeneity of the definitions used for TTD, some of which may not be adapted to the disease setting. There is a compelling need to standardize the TTD approach adapted to the cancer site and setting.

Acknowledgements This work was supported by a grant from the "Institut National du Cancer (INCA 11862)". The authors thank François Calais (University of Franche-Comté, Besançon, France) for developing and executing the search strategy. The authors thank Fiona Ecarnot (EA3920, University Hospital Besançon, University of Franche-Comté, Besançon, France) for editorial assistance.

Funding This work was funded by the "Institut National du Cancer (Grant INCA 11862)".

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Efficace, F., Fayers, P., Pusic, A., Cemal, Y., Yanagawa, J., Jacobs, M., et al. (2015). Quality of patient-reported outcome reporting across cancer randomized controlled trials according to the CONSORT patient-reported outcome extension: A pooled analysis of 557 trials. *Cancer, 121*(18), 3335–3342. https://doi.org/10.1002/cncr.29489.
2. Secord, A. A., Coleman, R. L., Havrilesky, L. J., Abernethy, A. P., Samsa, G. P., & Cella, D. (2015). Patient-reported outcomes as end points and outcome indicators in solid tumours. *Nature Reviews. Clinical Oncology, 12*(6), 358–370. https://doi.org/10.1038/nrclinonc.2015.29.
3. Bonnetain, F., Fiteni, F., Efficace, F., & Anota, A. (2016). Statistical challenges in the analysis of health-related quality of life in cancer clinical trials. *Journal of Clinical Oncology, 34*(16), 1953–1956. https://doi.org/10.1200/JCO.2014.56.7974.
4. Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine, 16*(20), 2349–2380.
5. Fairclough, D. L. (2010). *Design and analysis of quality of life studies in clinical trials*. London: Chapman and Hall/CRC.
6. Anota, A., Hamidou, Z., Paget-Bailly, S., Chibaudel, B., Bascoul-Mollevi, C., Auquier, P., et al. (2015). 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? *Quality of Life Research, 24*(1), 5–18. https://doi.org/10.1007/s11136-013-0583-6.
7. Bonnetain, F., Dahan, L., Maillard, E., Ychou, M., Mitry, E., Hammel, P., et al. (2010). Time until definitive quality of life score deterioration as a means of longitudinal analysis for treatment trials in patients with metastatic pancreatic adenocarcinoma. *European Journal of Cancer, 46*(15), 2753–2762. https://doi.org/10.1016/j.ejca.2010.07.023.
8. Hamidou, Z., Dabakuyo, T. S., Mercier, M., Fraisse, J., Causeret, S., Tixier, H., et al. (2011). Time to deterioration in quality of life score as a modality of longitudinal analysis in patients with breast cancer. *The Oncologist, 16*(10), 1458–1468. https://doi.org/10.1634/theoncologist.2011-0085.
9. Cortes, J., Hudgens, S., Twelves, C., Perez, E. A., Awada, A., Yelle, L., et al. (2015). Health-related quality of life in patients with locally advanced or metastatic breast cancer treated with eribulin mesylate or capecitabine in an open-label randomized

phase 3 trial. *Breast Cancer Research and Treatment, 154*(3), 509–520. https://doi.org/10.1007/s10549-015-3633-7.
10. Verma, S., O'Shaughnessy, J., Burris, H. A., Campone, M., Alba, E., Chandiwana, D., et al. (2018). Health-related quality of life of postmenopausal women with hormone receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer treated with ribociclib + letrozole: Results from MONALEESA-2. *Breast Cancer Research and Treatment, 170*(3), 535–545. https://doi.org/10.1007/s10549-018-4769-z.
11. Mathoulin-Pelissier, S., Gourgou-Bourgade, S., Bonnetain, F., & Kramar, A. (2008). Survival end point reporting in randomized cancer clinical trials: A review of major journals. *Journal of Clinical Oncology, 26*(22), 3721–3726. https://doi.org/10.1200/JCO.2007.14.1192.
12. Higgins, J., & Green, S. (2014). Cochrane handbook for systematic reviews of interventions Version 5.1. 0 [updated March 2011]. 2011. *The Cochrane Collaboration*.
13. Al-Batran, S. E., Van Cutsem, E., Oh, S. C., Bodoky, G., Shimada, Y., Hironaka, S., et al. (2016). Quality-of-life and performance status results from the phase III RAINBOW study of ramucirumab plus paclitaxel versus placebo plus paclitaxel in patients with previously treated gastric or gastroesophageal junction adenocarcinoma. *Annals of Oncology, 27*(4), 673–679. https://doi.org/10.1093/annonc/mdv625.
14. Bang, Y.-J., Xu, R.-H., Chin, K., Lee, K.-W., Park, S. H., Rha, S. Y., et al. (2017). Olaparib in combination with paclitaxel in patients with advanced gastric cancer who have progressed following first-line therapy (GOLD): A double-blind, randomised, placebo-controlled, phase 3 trial. *The Lancet Oncology, 18*(12), 1637–1651. https://doi.org/10.1016/S1470-2045(17)30682-4.
15. Bascoul-Mollevi, C., Gourgou, S., Galais, M. P., Raoul, J. L., Bouche, O., Douillard, J. Y., et al. (2017). Health-related quality of life results from the PRODIGE 5/ACCORD 17 randomised trial of FOLFOX versus fluorouracil-cisplatin regimen in oesophageal cancer. *European Journal of Cancer, 84,* 239–249. https://doi.org/10.1016/j.ejca.2017.07.038.
16. Bodoky, G., Scheulen, M. E., Rivera, F., Jassem, J., Carrato, A., Moiseyenko, V., et al. (2015). Clinical benefit and health-related quality of life assessment in patients treated with cisplatin/S-1 versus cisplatin/5-FU: Secondary end point results from the First-Line Advanced Gastric Cancer Study (FLAGS). *Journal of Gastrointestinal Cancer, 46*(2), 109–117. https://doi.org/10.1007/s12029-014-9680-1.
17. Boye, M., Wang, X., Srimuninnimit, V., Kang, J. H., Tsai, C.-M., Orlando, M., et al. (2016). First-line pemetrexed plus cisplatin followed by gefitinib maintenance therapy versus gefitinib monotherapy in East Asian never-smoker patients with locally advanced or metastatic nonsquamous non-small-cell lung cancer: Quality of life results from a randomized phase III trial. *Clinical Lung Cancer, 17*(2), 150–160. https://doi.org/10.1016/j.cllc.2015.12.004.
18. Brahmer, J. R., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csőszi, T., Fülöp, A., et al. (2017). 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. *The Lancet Oncology, 18*(12), 1600–1609. https://doi.org/10.1016/S1470-2045(17)30690-3.
19. Cella, D., Escudier, B., Tannir, N. M., Powles, T., Donskov, F., Peltola, K., et al. (2018). Quality of life outcomes for cabozantinib versus everolimus in patients with metastatic renal cell carcinoma: METEOR phase III randomized trial. *Journal of Clinical Oncology, 36*(8), 757–764. https://doi.org/10.1200/JCO.2017.75.2170.
20. Chi, K. N., Protheroe, A., Rodríguez-Antolín, A., Facchini, G., Suttman, H., Matsubara, N., et al. (2018). Patient-reported outcomes following abiraterone acetate plus prednisone added to androgen deprivation therapy in patients with newly diagnosed metastatic castration-naive prostate cancer (LATITUDE): An international, randomised phase 3 trial. *The Lancet Oncology, 19*(2), 194–206. https://doi.org/10.1016/S1470-2045(17)30911-7.
21. Ellis, P. M., Shepherd, F. A., Millward, M., Perrone, F., Seymour, L., Liu, G., et al. (2014). Dacomitinib compared with placebo in pretreated patients with advanced or metastatic non-small-cell lung cancer (NCIC CTG BR.26): A double-blind, randomised, phase 3 trial. *The Lancet Oncology, 15*(12), 1379–1388. https://doi.org/10.1016/s1470-2045(14)70472-3.
22. Felip, E., Hirsh, V., Popat, S., Cobo, M., Fülöp, A., Dayen, C., et al. (2018). Symptom and quality of life improvement in LUX-Lung 8, an open-label phase III study of second-line afatinib versus erlotinib in patients with advanced Squamous Cell carcinoma of the lung after first-line platinum-based chemotherapy. *Clinical Lung Cancer, 19*(1), 74–83.e11. https://doi.org/10.1016/j.cllc.2017.06.002.
23. Fiteni, F., Anota, A., Bonnetain, F., Oster, J.-P., Pichon, E., Wislez, M., et al. (2016). Health-related quality of life in elderly patients with advanced non-small cell lung cancer comparing carboplatin and weekly paclitaxel doublet chemotherapy with monotherapy. *The European Respiratory Journal, 48*(3), 861–872. https://doi.org/10.1183/13993003.01695-2015.
24. Fizazi, K., Scher, H. I., Miller, K., Basch, E., Sternberg, C. N., Cella, D., et al. (2014). Effect of enzalutamide on time to first skeletal-related event, pain, and quality of life in men with castration-resistant prostate cancer: Results from the randomised, phase 3 AFFIRM trial. *The Lancet Oncology, 15*(10), 1147–1156. https://doi.org/10.1016/S1470-2045(14)70303-1.
25. Geater, S. L., Xu, C. R., Zhou, C., Hu, C. P., Feng, J., Lu, S., et al. (2015). Symptom and quality of life improvement in LUX-Lung 6: An open-label phase III study of afatinib versus cisplatin/gemcitabine in asian patients with EGFR mutation-positive advanced non-small-cell lung cancer. *Journal of Thoracic Oncology, 10*(6), 883–889. https://doi.org/10.1097/JTO.0000000000000517.
26. Gill, S., Ko, Y. J., Cripps, C., Beaudoin, A., Dhesy-Thind, S., Zulfiqar, M., et al. (2016). PANCREOX: A randomized phase III study of fluorouracil/leucovorin with or without oxaliplatin for second-line advanced pancreatic cancer in patients who have received gemcitabine-based chemotherapy. *Journal of Clinical Oncology, 34*(32), 3914–3920. https://doi.org/10.1200/JCO.2016.68.5776.
27. Guimbaud, R., Louvet, C., Ries, P., Ychou, M., Maillard, E., Andre, T., et al. (2014). Prospective, randomized, multicenter, phase III study of fluorouracil, leucovorin, and irinotecan versus epirubicin, cisplatin, and capecitabine in advanced gastric adenocarcinoma: A French intergroup (Federation Francophone de Cancerologie Digestive, Federation Nationale des Centres de Lutte Contre le Cancer, and Groupe Cooperateur Multidisciplinaire en Oncologie) study. *Journal of Clinical Oncology, 32*(31), 3520–3526. https://doi.org/10.1200/JCO.2013.54.1011.
28. Hagiwara, Y., Ohashi, Y., Okusaka, T., Ueno, H., Ioka, T., Boku, N., et al. (2017). Health-related quality of life in a randomised phase III study of gemcitabine plus S-1, S-1 alone and gemcitabine alone for locally advanced or metastatic pancreatic cancer: GEST study. *ESMO Open, 2*(1), e000151. https://doi.org/10.1136/esmoopen-2016-000151.
29. Hamidou, Z., Chibaudel, B., Hebbar, M., Hug de Larauze, M., André, T., Louvet, C., et al. (2016). Time to definitive health-related quality of life score deterioration in patients with resectable metastatic colorectal cancer treated with FOLFOX4 versus sequential dose-dense FOLFOX7 followed by FOLFIRI: The MIROX randomized phase III trial. *PLoS ONE, 11*(6), e0157067. https://doi.org/10.1371/journal.pone.0157067.
30. Harbeck, N., Iyer, S., Turner, N., Cristofanilli, M., Ro, J., Andre, F., et al. (2016). Quality of life with palbociclib plus fulvestrant

in previously treated hormone receptor-positive, HER2-negative metastatic breast cancer: Patient-reported outcomes from the PALOMA-3 trial. *Annals of Oncology, 27*(6), 1047–1054. https://doi.org/10.1093/annonc/mdw139.

31. Harrington, K. J., Ferris, R. L., Blumenschein, G., Colevas, A. D., Fayette, J., Licitra, L., et al. (2017). Nivolumab versus standard, single-agent therapy of investigator's choice in recurrent or metastatic squamous cell carcinoma of the head and neck (CheckMate 141): Health-related quality-of-life results from a randomised, phase 3 trial. *The Lancet Oncology, 18*(8), 1104–1115. https://doi.org/10.1016/S1470-2045(17)30421-7.
32. Lee, C. K., Novello, S., Rydén, A., Mann, H., & Mok, T. (2018). Patient-reported symptoms and impact of treatment with osimertinib versus chemotherapy in advanced non-small-cell lung cancer: The AURA3 trial. *Journal of Clinical Oncology, 36*(18), 1853–1860. https://doi.org/10.1200/JCO.2017.77.2293.
33. Long, G. V., Atkinson, V., Ascierto, P. A., Robert, C., Hassel, J. C., Rutkowski, P., et al. (2016). Effect of nivolumab on health-related quality of life in patients with treatment-naive advanced melanoma: Results from the phase III CheckMate 066 study. *Annals of Oncology, 27*(10), 1940–1946. https://doi.org/10.1093/annonc/mdw265.
34. Loriot, Y., Miller, K., Sternberg, C. N., Fizazi, K., De Bono, J. S., Chowdhury, S., et al. (2015). Effect of enzalutamide on health-related quality of life, pain, and skeletal-related events in asymptomatic and minimally symptomatic, chemotherapy-naive patients with metastatic castration-resistant prostate cancer (PREVAIL): Results from a randomised, phase 3 trial. *The Lancet Oncology, 16*(5), 509–521. https://doi.org/10.1016/S1470-2045(15)70113-0.
35. Machiels, J.-P. H., Haddad, R. I., Fayette, J., Licitra, L. F., Tahara, M., Vermorken, J. B., et al. (2015). Afatinib versus methotrexate as second-line treatment in patients with recurrent or metastatic squamous-cell carcinoma of the head and neck progressing on or after platinum-based therapy (LUX-Head & Neck 1): An open-label, randomised phase 3 trial. *The Lancet Oncology, 16*(5), 583–594. https://doi.org/10.1016/S1470-2045(15)70124-5.
36. Taphoorn, M. J., Dirven, L., Kanner, A. A., Lavy-Shahaf, G., Weinberg, U., Taillibert, S., et al. (2018). Influence of treatment with tumor-treating fields on health-related quality of life of patients with newly diagnosed glioblastoma: A secondary analysis of a randomized clinical trial. *JAMA Oncology, 4*(4), 495–504. https://doi.org/10.1001/jamaoncol.2017.5082.
37. Novello, S., Kaiser, R., Mellemgaard, A., Douillard, J. Y., Orlov, S., Krzakowski, M., et al. (2015). Analysis of patient-reported outcomes from the LUME-Lung 1 trial: A randomised, double-blind, placebo-controlled, Phase III study of second-line nintedanib in patients with advanced non-small cell lung cancer. *European Journal of Cancer, 51*(3), 317–326. https://doi.org/10.1016/j.ejca.2014.11.015.
38. Pavel, M. E., Singh, S., Strosberg, J. R., Bubuteishvili-Pacaud, L., Degtyarev, E., Neary, M. P., et al. (2017). Health-related quality of life for everolimus versus placebo in patients with advanced, non-functional, well-differentiated gastrointestinal or lung neuroendocrine tumours (RADIANT-4): A multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. *The Lancet Oncology, 18*(10), 1411–1422. https://doi.org/10.1016/S1470-2045(17)30471-0.
39. Perol, M., Ciuleanu, T. E., Arrieta, O., Prabhash, K., Syrigos, K. N., Goksel, T., et al. (2016). Quality of life results from the phase 3 REVEL randomized clinical trial of ramucirumab-plus-docetaxel versus placebo-plus-docetaxel in advanced/metastatic non-small cell lung cancer patients with progression after platinum-based chemotherapy. *Lung Cancer, 93,* 95–103. https://doi.org/10.1016/j.lungcan.2016.01.007.
40. Petrella, T. M., Robert, C., Richtig, E., Miller, W. H., Jr., Masucci, G. V., Walpole, E., et al. (2017). Patient-reported outcomes in KEYNOTE-006, a randomised study of pembrolizumab versus ipilimumab in patients with advanced melanoma. *European Journal of Cancer, 86,* 115–124. https://doi.org/10.1016/j.ejca.2017.08.032.
41. Qin, S., Bi, F., Jin, J., Cheng, Y., Guo, J., Ren, X., et al. (2015). Axitinib versus sorafenib as a second-line therapy in Asian patients with metastatic renal cell carcinoma: Results from a randomized registrational study. *OncoTargets and Therapy, 8,* 1363–1373. https://doi.org/10.2147/OTT.S83302.
42. Reck, M., Taylor, F., Penrod, J. R., DeRosa, M., Morrissey, L., Dastani, H., et al. (2018). Impact of nivolumab versus docetaxel on health-related quality of life and symptoms in patients with advanced squamous non-small cell lung cancer: results from the CheckMate 017 study. *Journal of Thoracic Oncology, 13*(2), 194–204. https://doi.org/10.1016/j.jtho.2017.10.029.
43. Saad, F., Fizazi, K., Jinga, V., Efstathiou, E., Fong, P. C., Hart, L. L., et al. (2015). Orteronel plus prednisone in patients with chemotherapy-naive metastatic castration-resistant prostate cancer (ELM-PC 4): A double-blind, multicentre, phase 3, randomised, placebo-controlled trial. *The Lancet Oncology, 16*(3), 338–348. https://doi.org/10.1016/S1470-2045(15)70027-6.
44. Satoh, T., Bang, Y.-J., Gotovkin, E. A., Hamamoto, Y., Kang, Y.-K., Moiseyenko, V. M., et al. (2014). Quality of life in the trastuzumab for gastric cancer trial. *The Oncologist, 19*(7), 712–719. https://doi.org/10.1634/theoncologist.2014-0058.
45. Shiroiwa, T., Fukuda, T., Shimozuma, K., Mouri, M., Hagiwara, Y., Doihara, H., et al. (2017). Long-term health status as measured by EQ-5D among patients with metastatic breast cancer: Comparison of first-line oral S-1 and taxane therapies in the randomized phase III SELECT BC trial. *Quality of Life Research, 26*(2), 445–453. https://doi.org/10.1007/s11136-016-1388-1.
46. Taphoorn, M. J., Henriksson, R., Bottomley, A., Cloughesy, T., Wick, W., Mason, W. P., et al. (2015). Health-related quality of life in a randomized phase III study of bevacizumab, temozolomide, and radiotherapy in newly diagnosed glioblastoma. *Journal of Clinical Oncology, 33*(19), 2166–2175. https://doi.org/10.1200/JCO.2014.60.3217.
47. Vaughn, D. J., Bellmunt, J., Fradet, Y., Lee, J. L., Fong, L., Vogelzang, N. J., et al. (2018). Health-related quality-of-life analysis from KEYNOTE-045: A phase III study of pembrolizumab versus chemotherapy for previously treated advanced urothelial cancer. *Journal of Clinical Oncology, 36*(16), 1579–1587. https://doi.org/10.1200/JCO.2017.76.9562.
48. Vinik, A., Bottomley, A., Korytowsky, B., Bang, Y.-J., Raoul, J.-L., Valle, J. W., et al. (2016). Patient-reported outcomes and quality of life with sunitinib versus placebo for pancreatic neuroendocrine tumors: Results from an international phase III trial. *Targeted Oncology, 11*(6), 815–824. https://doi.org/10.1007/s11523-016-0462-5.
49. Yang, Y.-P., Ma, Y.-X., Huang, Y., Zhao, Y.-Y., Fang, W.-F., Hong, S.-D., et al. (2015). QoL analyses from INFORM study, a phase III study of gefitinib versus placebo as maintenance therapy in advanced NSCLC. *Scientific Reports, 5,* 11934. https://doi.org/10.1038/srep11934.
50. Fiteni, F., Westeel, V., Pivot, X., Borg, C., Vernerey, D., & Bonnetain, F. (2014). Endpoints in cancer clinical trials. *Journal of Visceral Surgery, 151*(1), 17–22.
51. Montori, V. M., Permanyer-Miralda, G., Ferreira-González, I., Busse, J. W., Pacheco-Huergo, V., Bryant, D., et al. (2005). Validity of composite end points in clinical trials. *BMJ (Clinical Research ed.), 330*(7491), 594–596. https://doi.org/10.1136/bmj.330.7491.594.
52. Freemantle, N., Calvert, M., Wood, J., Eastaugh, J., & Griffin, C. (2003). Composite outcomes in randomized trials: Greater precision but with greater uncertainty? *JAMA, 289*(19), 2554–2559. https://doi.org/10.1001/jama.289.19.2554.

53. Anota, A., Savina, M., Bascoul-Mollevi, C., & Bonnetain, F. (2017). Qolr: An r package for the longitudinal analysis of health-related quality of life in oncology. *Journal of Statistical Software, Articles, 77*(12), 1–30.
54. Bottomley, A., Pe, M., Sloan, J., Basch, E., Bonnetain, F., Calvert, M., et al. (2016). Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: A start in setting international standards. *The Lancet Oncology, 17*(11), e510–e514. https://doi.org/10.1016/S1470-2045(16)30510-1.
55. Ajani, J. A., Moiseyenko, V. M., Tjulandin, S., Majlis, A., Constenla, M., Boni, C., et al. (2007). Quality of life with docetaxel plus cisplatin and fluorouracil compared with cisplatin and fluorouracil from a phase III trial for advanced gastric or gastroesophageal adenocarcinoma: The V-325 Study Group. *Journal of Clinical Oncology, 25*(22), 3210–3216. https://doi.org/10.1200/JCO.2006.08.3956.
56. Panageas, K. S., Ben-Porat, L., Dickler, M. N., Chapman, P. B., & Schrag, D. (2007). When you look matters: The effect of assessment schedule on progression-free survival. *Journal of the National Cancer Institute, 99*(6), 428–432. https://doi.org/10.1093/jnci/djk091.
57. Kiebert, G. M., Curran, D., & Aaronson, N. K. (1998). Quality of life as an endpoint in EORTC clinical trials. European Organization for Research and Treatment for Cancer. *Statistics in Medicine, 17*(5–7), 561–569.
58. Fiteni, F., Pam, A., Anota, A., Vernerey, D., Paget-Bailly, S., Westeel, V., et al. (2015). Health-related quality-of-life as co-primary endpoint in randomized clinical trials in oncology. *Expert Review of Anticancer Therapy, 15*(8), 885–891.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2. Développement de deux macro-programmes SAS : %TTD et %TUDD

Article : %TTD and %TUDD: new SAS macro programs to calculate the survival data of the time to deterioration for patient-reported outcomes data in oncology.

Article en cours de relecture

Résumé

Objectif

L'analyse longitudinale des données de PROs reste problématique, car aucune standardisation des méthodes statistiques n'a réellement été proposée. Cela compromet la comparaison des résultats entre les différents essais cliniques. Dans ce contexte, l'approche du temps jusqu'à détérioration a récemment été proposée et est régulièrement utilisée comme modalité d'analyse longitudinale des données de PROs en cancérologie.

Méthodes

Deux nouveaux macro-programmes SAS ont été développés, %TTD et %TUDD, permettant de réaliser l'analyse longitudinale des données de PROs selon l'approche du temps jusqu'à détérioration. L'ensemble des définitions de détérioration recommandées à ce jour a été implémenté.

Résultats

Nous avons décrit les programmes et leurs différentes fonctionnalités. La macro %TTD permet d'estimer le temps jusqu'à une première détérioration ou une détérioration transitoire. La macro %TUDD estime, quant à elle, le temps jusqu'à une détérioration définitive. Nous illustrons nos programmes en présentant différentes applications sur l'essai clinique randomisé de phase II AFUGEM GERCOR.

Conclusion

L'implémentation des définitions du temps jusqu'à détérioration sous le logiciel SAS facilitera l'application de cette approche auprès de la communauté scientifique. Elle permettra la progression vers l'objectif de standardisation de l'analyse longitudinale des données de PROs dans les essais cliniques en cancérologie.

%TTD and %TUDD: new SAS macro programs to calculate the survival data of the time to deterioration for patient-reported outcomes data in oncology

E. Charton[*,1,2,3], A. Falcoz[1,2], E. François[4], C. Touraine[5], J.-B. Bachet[6], C. Louvet[7], Z. Hamidou[3,8], C. Bascoul-Mollevi[3,5,9], A. Anota[1,2,3]

[1]Methodology and Quality of Life Unit in Oncology, University Hospital of Besançon, Besançon, France

[2]University Bourgogne Franche-Comté, INSERM, EFS BFC, UMR1098, Interactions Hôte-Greffon-Tumeur/Ingénierie Cellulaire et Génique, France

[3]French National Platform Quality of Life and Cancer, Besançon, France

[4]IVIDATA Group, Paris, France

[5]Montpellier Cancer Institute (ICM), University of Montpellier, Montpellier, France

[6]Department of Hepato-Gastroenterology, Groupe hospitalier Pitié Salpêtrière, Sorbonne University, UPMC University, Paris, France

[7]Department of Oncology, Institut Mutualiste Montsouris, Paris, France

[8]EA3279 Self-perceived Health Assessment Research Unit, Aix-Marseille University, Marseille, France

[9]Institute of Cancer Research of Montpellier (IRCM), ICM, INSERM, Montpellier, France

*Corresponding author: Ms Emilie Charton, Methodology and Quality of Life Unit in Oncology (INSERM UMR 1098), University Hospital of Besançon, 3 Boulevard Fleming, 25030 Besançon, France. Tel: +33 3 70 63 21 87; E-mail: echarton@chu-besancon.fr

ABSTRACT

Background and objective: Longitudinal analysis of patient-reported outcome (PRO) data remains challenging, as no standardization of statistical methods has actually been proposed, making comparison of PRO results between clinical trials difficult. In this context, the time to deterioration approach has recently been proposed and is regularly used as a modality of longitudinal PRO analysis in oncology.

Methods: Two new SAS macro programs were developed, `%TTD` and `%TUDD`, which implement longitudinal analysis of PRO data according to the time to deterioration approach. These programs implement the recommended deterioration definitions. We described the programs with their different functionalities.

Results: The `%TTD` macro allows the calculation of the time to first or transient deterioration, and the `%TUDD` macro calculates the time until definitive deterioration. We illustrate our programs by presenting different applications on the randomized phase II AFUGEM GERCOR clinical trial.

Conclusion: The implementation of the deterioration definitions in SAS software allows the dissemination of this approach, in order to move toward the goal of standardization of longitudinal PRO analysis in oncology clinical trials.

Keywords: Patient-reported outcomes; Longitudinal analysis; Time to deterioration; Oncology; SAS macro

Highlights

- The new SAS macro programs, `%TTD` and `%TUDD`, estimate the time to deterioration.
- Several definitions of time to deterioration were implemented following the recent recommendations published.
- To our knowledge, these are the first SAS macros implementing this method.
- These macros help to move toward the goal of standardization.

1. Introduction

In oncology clinical trials, patient-reported outcomes (PROs), including health-related quality of life (HRQoL), are important endpoints to assess the clinical benefit of new treatment strategies for the patient [1,2].

PROs are generally assessed using cancer-specific questionnaires, the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire – Core 30 (EORTC QLQ-C30) [3] or the Functional Assessment of Cancer Therapy – General [4]. To illustrate, the QLQ-C30 includes 30 items and assesses a global health status, 5 functional scales (physical, role, emotional, cognitive and social functioning) and 9 symptomatic scales (fatigue, nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). Scores are generated according to the EORTC scoring manual [5]. Scores vary from 0 (worst) to 100 (best) for global health status and functional scales, and from 0 (best) to 100 (worst) for the symptom scales.

The PROs questionnaires are fulfilled at several times during the treatment, allowing to assess its impact on patients' PRO level over time. In order to account for the correlation between the repeated PRO measures on the same patient, a longitudinal analysis is recommended. However, the longitudinal analysis of PRO data remains challenging, including the choice of the statistical method, the way of handling the missing data and the interpretation of the results [6]. In particular, the PROs should be interpreted in term of clinical relevance using the minimal important difference (MID). The MID was defined as the smallest difference in score in the outcome of interest that patients perceive as important [7].

Several methods have been proposed for the longitudinal analysis of PRO data including the linear mixed model [8,9] and the time to deterioration approach [10]. This latter method has recently been proposed and is regularly used in phase III clinical trials [11]. Some initial recommendations have previously been proposed, emphasizing the need to adapt the definition based on the therapeutic setting [10]. More recently, the Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints Data (SISAQOL) consortium proposed new standardized

deterioration definitions [12]. The time to deterioration approach is attractive to many clinicians since it offers results that are easily understandable and interpretable by the methodology used.

The objective of this work is to propose new SAS macro programs. Two SAS macros, called `%TTD` and `%TUDD`, were developed to determine the longitudinal analysis of PRO data according to the time to deterioration approach.

2. Material and methods

Several definitions used for the time to deterioration have been proposed depending on the definition of what is considered to be a « deterioration ». This depends on the reference score, the MID used to qualify the deterioration, as well as on censoring rules. The choice of the reference score to qualify the deterioration is a major concern. Generally, the reference score is the baseline score (i.e. before randomization or at inclusion in the absence of randomization). However, the patients' adaptation process to the disease and treatment may mean that patients do not necessarily assess their HRQoL with the same standards, reflecting a response shift effect [13]. Thus, the baseline score may not be systematically the reference score. Different reference scores have been proposed to explore the potential impact of a response shift such as the best previous score (i.e. the best PRO level already experienced by the patient) and the immediately preceding score. The time to deterioration method can take into account the occurrence of the response shift effect recalibration component by choosing different reference scores to qualify the deterioration.

The deterioration could be considered as « reversible », « transient » or « definitive ». Composite events may also be considered, for example, including death in the event definition. The time to first clinically significant deterioration (TTD) has been defined as the time between patient inclusion into the study and a first deterioration of at least one MID unit as compared to the reference score (« reversible » deterioration) [10]. Patients with no deterioration before their dropout from the study are censored at the time of the last PRO assessment. The « transient » deterioration definition is close to

the TTD definition, but after the post-baseline deterioration, at least one follow-up score should be higher than or at least be at the baseline level. Finally, the « definitive » deterioration corresponds to the time until definitive deterioration (TUDD) definition, which has been defined in different ways as follows:

- the time between patient inclusion into the study and a first deterioration of at least one MID unit as compared to the reference score, without further improvement of more than one MID unit as compared to the reference score or without any available score;
- the time between patient inclusion into the study and a first deterioration of at least one MID unit as compared to the reference score, thereafter maintaining this deterioration of at least one MID unit for all following scores (i.e. the deterioration is observed for all the scores following the first deterioration) or without any available score;
- the time between patient inclusion into the study and a first deterioration if at least one MID unit as compared to the reference score, with no further improvement of more than one MID unit as compared to the score qualifying the deterioration (i.e. the score at the time of the first deterioration observe) or without any available score;
- the time between patient inclusion into the study and a first deterioration if at least one MID unit as compared to the reference score, with no further improvement of more than the score qualifying the deterioration or without any available score.

Intermittent missing data are generally ignored in the time to deterioration approach, supposing that the PRO level remains unchanged between two available PRO measurements.

Both SAS macro programs `%TTD` and `%TUDD` implement all these published deterioration definitions [10,12]. A summary of the event definitions implemented in the macro programs is presented in *Table 1*.

3. SAS macro programs description

Both SAS macro programs compute the different definitions of deterioration previously presented. Both macros requires prior a SAS dataset (`dataset`) with the following variables: (1) the patient identification number variable (`npat`); (2) the time identification variable (`time`) where baseline measure must be equal to 0; (3) the date of PRO measurement variable (`date`); (4) the PRO score(s) variable(s) (`score`) and (5) the date of death variable (`date_death`), only when death is considered as an event. If death is considered as an event, the date of death variable must contain the date of death for patients who died during the study and missing values for patients still alive at the end of the study. This dataset and its parameters are needed to ensure the efficient running of macros. The dataset must also be sorted by patient identification number and time identification. The missing values must be represented by a point. Thus, depending on the deterioration definition chosen, the user must specify the following macro variables:

- The name of the SAS dataset in the format presented above (`dataset`);
- The name of the patient identification number variable in the dataset (`npat`);
- The name of the time identification variable in the dataset (`time`);
- The name of the date of PRO measurement variable in the dataset (`date`);
- The list of the name(s) of the PRO score(s) variable(s) in the dataset to be studied (`score`). Multiple PRO scores can be treated for one application of macro, using the « | » command as a separator (e.g. `score=QL|SL`, corresponding to the global health status (QL) and insomnia (SL) score from the QLQ-C30);
- The list of threshold(s) for the MID considered (`MID`). If *x* PRO scores are considered in the `score` option, then *x* `MID` must be considered using the « | » command as a separator (e.g. if `score=QL|SL` then `MID=5|5`);
- The list of value(s) of the `order` option: whether the deterioration corresponds to a decrease (`order=1`) or an increase of the PRO score (`order=2`). If *x* PRO scores are considered in

the `score` option, then *x* `order` must be considered using the « | » command as a separator (e.g. if `score=QL|SL` then `order=1|2`);

- The reference score used to qualify the deterioration (`ref`): the baseline score (`ref=0`), the best previous score (`ref=1`) or the immediately preceding score (`ref=2`);
- The consideration of the baseline missing scores (`no_baseline`): whether patients with no baseline score are censored at baseline (`no_baseline=0`), deteriorated at baseline (`no_baseline=1`) or excluded of the analysis (`no_baseline=2`);
- The consideration of the no follow-up score (`no_follow`): whether patients with no follow-up score are censored (`no_follow=0`) or deteriorated (`no_follow=1`) one day after baseline;
- How death is taken into account (`death`): whether patients who died during the study with no deterioration observed before death are censored at the time of the last available PRO measurement (`death=0`) or all-cause death are considered as an event at date of death (`death=1`);
- The name of the variable in the dataset containing the date of death (`date_death`) only when death is considered as an event. If death is censored, then `date_death=NA`;
- The name of the output dataset (`out`).

All these macro variables are present in both macros. The SAS codes for each macro are provided in *Supplementary Tables A* and *B* respectively.

3.1. `%TTD` macro

The `%TTD` macro estimates the time to first or transient deterioration. The syntax is the following: `%TTD(dataset=, npat=, time=, date=, score=, MID=, order=, ref=, definition=, no_baseline=, no_follow=, death=, date_death=, out=).`

The full description of the macro variables that must have been entered by the user are listed in *Table*

2. One supplementary macro variable is required, namely `definition`, in which the notion of « deterioration » needs to be specified, according to the two definitions proposed in the section 2:

- a first deterioration (`definition=1`), i.e. a first deterioration of at least one MID unit as compared to the reference score;
- a transient deterioration (`definition=2`), i.e. a first deterioration of at least one MID unit as compared to the reference score and after the post-baseline deterioration at least one follow-up score higher than or at least be at the baseline level.

3.2. `%TUDD` macro

The `%TUDD` macro estimates the time until definitive deterioration. The syntax is the following: `%TUDD(dataset=, npat=, time=, date=, score=, MID=, order=, ref=, definitive=, no_baseline=, no_follow=, death=, date_death=, out=).`

This syntax is almost similar to `%TTD`. *Table 3* describes the macro variables that must have been entered by the user. One supplementary macro variable is required, namely `definitive`, in which the notion of « definitive deterioration » needs to be specified, according to the four definitions proposed in the section 2:

- with no further improvement of more than one MID unit as compared to the reference score (`definitive=1`);
- maintaining this deterioration of at least one MID unit for all following scores (`definitive=2`), i.e. the deterioration is observed for all the scores following the first deterioration;
- with no further improvement of more than one MID unit as compared to the score qualifying the deterioration (`definitive=3`);
- with no further improvement of more than the score qualifying the deterioration (`definitive=4`).

4. Example

In this section we illustrate an application of each macro to estimate the time to deterioration for PROs data from the randomized phase II AFUGEM GERCOR clinical trial [14]. The detailed of the study design, the administration and the compliance of the QLQ-C30 have previously been reported [15,16]. Five targeted dimensions were defined a priori in the protocol, namely: physical functioning (PF), emotional functioning (EF), fatigue (FA), pain (PA), and appetite loss (AP).

The `%TTD` and `%TUDD` macros were applied on these five targeted dimensions. The *Supplementary Table C* provides the HRQoL scores of some patients from the AFUGEM study used as the SAS dataset, called « QoL_scores » (`dataset=QoL_scores`), to apply the macros.

We used the `%TTD` macro to estimate the time to transient deterioration (`definition=2`) for the five targeted dimensions (`score=PF|EF|FA|PA|AP`) with a 5-point MID (`MID=5|5|5|5|5`) as compared to the baseline score (`ref=0`). A deterioration for the physical and emotional functioning corresponds to a decrease of the score, while deterioration for the other three symptomatic scores corresponds to an increase of the score (`order=1|1|2|2|2`). Patients with no baseline score are excluded of the analysis (`no_baseline=2`) and patients with a baseline score but no follow-up score are censored one day after baseline (`no_follow=0`). The death was not considered as an event, thus patients who died during the study with no deterioration before are censored at the time of the last available PRO measurement (`death=0` and `date_death=NA`). The *Table 4* presents the output dataset of the estimation obtained (`out=TTD_definition2`). The output dataset contains the patient identification (`npat`), the corresponding event and time in months of the five targeted dimensions.

Then, we applied the `%TUDD` macro to estimate the time until definitive deterioration (`definitive=2`) for the five targeted dimensions with a 5-point MID as compared to the baseline score (`ref=0`). Patients with no baseline score are censored at baseline (`no_baseline=0`) and patients with a baseline score but no follow-up score are censored one day after baseline

(`no_follow=0`). The death was not considered as an event (`death=0` and `date_death=NA`). The *Table 5* presents the output dataset of the estimation obtained (`out=TUDD_definitive2`). The output dataset contains the patient identification (`npat`), the corresponding event and time in months of the five targeted dimensions.

5. Discussion

This article presents new SAS macro programs, `%TTD` and `%TUDD`, for implementing the time to deterioration approach. The time to deterioration is a didactic and promising approach more and more used for the longitudinal analysis of PRO data in oncology clinical trials [11]. This approach requires a clear definition of what is considered to be a « deterioration ». Several initial definitions of the time to deterioration have already been proposed, based on the therapeutic setting [10] and revised by the SISAQoL consortium. In fact, new recommendations are being developed by the SISAQoL group [12]. These new recommendations do not take into consideration the cancer location or the therapeutic setting at the moment.

The SAS macro programs implement recommended published deterioration definitions and offer several options such as the choice of the reference score, the manner in which missing baseline or follow-up scores were considered and whether or not death is included as an event. Different MID can be considered by score. These macros provide an output dataset which contains the patient identification number, the event obtained (i.e. equals to one if the patient is deteriorated and 0 if not) and the delay of deterioration computed (i.e. a time variable equals to the time to deterioration or the time to censure). The variables event and time are created by score.

The implementation of the time to deterioration approach has also been implemented in R software [17] and Stata [18]. The R package allows also to compute the scores from the questionnaires developed by the EORTC group. The Stata command is dedicated to analysis the EORTC QLQ-C30. Cross validations were performed between software, in order to ensure the reliability of the results.

The programs have been tested on real and simulated datasets. The development of these different programs facilitates and allows standardization of the application of the time to deterioration approach under the three major statistical software.

In opposition to the other implemented programs, the `%TTD` and `%TUDD` macros require a previously calculation of the scores in order to ensure their efficient running. Indeed, the `order` parameter determines whether the deterioration corresponds to a decrease or an increase of the score. This requirement could be seen as a limitation of macros, but it permits these two macros to be applied to all PROs and not only to HRQoL questionnaires.

The two SAS macro programs provide automatic and reliable estimations of the time to deterioration approach. Thus, the different datasets obtained from the macros can be used to estimate deterioration curves using the Kaplan-Meier estimation method or the treatment effect using the Cox model.

6. Conclusion

To conclude, there is still a real compelling need to standardize the time to deterioration approach. The implementation of the time to deterioration definitions in SAS macros allows the wider dissemination in order to achieve the goal of standardization of longitudinal PRO analysis in oncology clinical trials.

Acknowledgments

This work was supported by a grant from the « Institut National du Cancer (INCA 11862) ».

References

1. Osoba, D. (2011). Health-related quality of life and cancer clinical trials. *Ther Adv Med Oncol, 3*(2), 57-71, doi:10.1177/1758834010395342.

2. Secord, A. A., Coleman, R. L., Havrilesky, L. J., Abernethy, A. P., Samsa, G. P., & Cella, D. (2015). Patient-reported outcomes as end points and outcome indicators in solid tumours. *Nat Rev Clin Oncol, 12*(6), 358-370, doi:10.1038/nrclinonc.2015.29.

3. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., et al. (1993). 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, 85*(5), 365-376, doi:10.1093/jnci/85.5.365.

4. Cella, D. F., Tulsky, D. S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., et al. (1993). The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol, 11*(3), 570-579, doi:10.1200/JCO.1993.11.3.570.

5. Fayers, P. M., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., Bottomley, A., et al. (2001). The EORTC QLQ-C30 Scoring Manual (3rd edition).

6. Bonnetain, F., Fiteni, F., Efficace, F., & Anota, A. (2016). Statistical Challenges in the Analysis of Health-Related Quality of Life in Cancer Clinical Trials. *J Clin Oncol, 34*(16), 1953-1956, doi:10.1200/JCO.2014.56.7974.

7. Jaeschke, R., Singer, J., & Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials, 10*(4), 407-415, doi:10.1016/0197-2456(89)90005-6.

8. Cnaan, A., Laird, N. M., & Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine, 16*(20), 2349-2380.

9. Fairclough, D. L. (2010). *Design and analysis of quality of life studies in clinical trials*: Chapman and Hall/CRC.

10. Anota, A., Hamidou, Z., Paget-Bailly, S., Chibaudel, B., Bascoul-Mollevi, C., Auquier, P., et al. (2015). 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? *Quality of Life Research, 24*(1), 5-18, doi:10.1007/s11136-013-0583-6.

11. Charton, E., Cuer, B., Cottone, F., Efficace, F., Touraine, C., Hamidou, Z., et al. (2020). Time to deterioration in cancer randomized clinical trials for patient-reported outcomes data: a systematic review. *Qual Life Res, 29*(4), 867-878, doi:10.1007/s11136-019-02367-7.

12. Coens, C., Pe, M., Dueck, A. C., Sloan, J., Basch, E., Calvert, M., et al. (2020). International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. *Lancet Oncol, 21*(2), e83-e96, doi:10.1016/S1470-2045(19)30790-9.

13. Sprangers, M. A., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine (1982), 48*(11), 1507-1515.

14. Bachet, J. B., Chibaudel, B., Bonnetain, F., Validire, P., Hammel, P., Andre, T., et al. (2015). A randomized phase II study of weekly nab-paclitaxel plus gemcitabine or simplified LV5FU2 as first-line therapy in patients with metastatic pancreatic cancer: the AFUGEM GERCOR trial. *BMC Cancer, 15*, 653, doi:10.1186/s12885-015-1656-4.

15. Bachet, J. B., Hammel, P., Desrame, J., Meurisse, A., Chibaudel, B., Andre, T., et al. (2017). Nab-paclitaxel plus either gemcitabine or simplified leucovorin and fluorouracil as first-line therapy for metastatic pancreatic adenocarcinoma (AFUGEM GERCOR): a non-comparative, multicentre, open-label, randomised phase 2 trial. *Lancet Gastroenterol Hepatol, 2*(5), 337-346, doi:10.1016/S2468-1253(17)30046-8.

16. Charton, E., Bachet, J. B., Hammel, P., Desrame, J., Chibaudel, B., Cohen, R., et al. (2019).

Impact on health-related quality of life deterioration-free survival of a first-line therapy combining nab-paclitaxel plus either gemcitabine or simplified leucovorin and fluorouracil for patients with metastatic pancreatic cancer: Results of the randomized phase II AFUGEM GERCOR clinical trial. *Cancer Med, 8*(11), 5079-5088, doi:10.1002/cam4.2311.

17. Anota, A., Savina, M., Bascoul-Mollevi, C., & Bonnetain, F. (2017). Qolr: An r package for the longitudinal analysis of health-related quality of life in oncology. *Journal of Statistical Software, Articles, 77*(12), 1-30.

18. Bascoul-Mollevi, C., Savina, M., Anota, A., Barbieri, A., Azria, D., Bonnetain, F., et al. (2018). Longitudinal health-related quality of life analysis in oncology with time to event approaches, the STATA command qlqc30_TTD. *Comput Methods Programs Biomed, 158*, 153-159, doi:10.1016/j.cmpb.2018.02.010.

Table 1. Summary of the event definitions implemented.

Definition	Reference score (X_{ref})				Event definition at time *i*	
	Baseline score	Best previous score		Immediately preceding score	Event decrease[1]	Event increase[2]
		Event decrease[1]	Event increase[2]			
TTD ≥ k points	X_0	$\max_{j<i}(X_j)$	$\min_{j<i}(X_j)$	X_{i-1}	$\exists i > 0, X_i \leq X_{ref} - k$	$\exists i > 0, X_i \geq X_{ref} + k$
TTD ≥ k points with at least one follow-up score higher than or at least be at the baseline level	X_0	–	–	–	$\exists i > 0, X_i \leq X_{ref} - k$ and $\exists j > i, X_j \geq X_{ref}$	$\exists i > 0, X_i \geq X_{ref} + k$ and $\exists j > i, X_j \leq X_{ref}$
TUDD ≥ k points with no further improvement > k points as compared to the reference score	X_0	$\max_{j<i}(X_j)$	$\min_{j<i}(X_j)$	X_{i-1}	$\exists i > 0, X_i \leq X_{ref} - k$ and $\forall j > i, X_j \leq X_{ref} + k$	$\exists i > 0, X_i \geq X_{ref} + k$ and $\forall j > i, X_j \geq X_{ref} - k$
TUDD ≥ k points with a deterioration observed for all scores following the first deterioration	X_0	$\max_{j<i}(X_j)$	$\min_{j<i}(X_j)$	X_{i-1}	$\exists i > 0, X_i \leq X_{ref} - k$ and $\forall j > i, X_j \leq X_{ref} - k$	$\exists i > 0, X_i \geq X_{ref} + k$ and $\forall j > i, X_j \geq X_{ref} + k$
TUDD ≥ k points with no further improvement > k points as compared to the score qualifying the deterioration	X_0	$\max_{j<i}(X_j)$	$\min_{j<i}(X_j)$	X_{i-1}	$\exists i > 0, X_i \leq X_{ref} - k$ and $\forall j > i, X_j \leq X_i + k$	$\exists i > 0, X_i \geq X_{ref} + k$ and $\forall j > i, X_j \geq X_i - k$
TUDD ≥ k points with no further improvement of more than the score qualifying the deterioration	X_0	$\max_{j<i}(X_j)$	$\min_{j<i}(X_j)$	X_{i-1}	$\exists i > 0, X_i \leq X_{ref} - k$ and $\forall j > i, X_j \leq X_i$	$\exists i > 0, X_i \geq X_{ref} + k$ and $\forall j > i, X_j \geq X_i$

TTD, time to first deterioration; TUDD, time until definitive deterioration; k, the minimal important difference; X, patient-reported outcome score; –, not recommended to use for this definition; [1]the deterioration corresponds to a decrease of the PRO score; [2]the deterioration corresponds to an increase of the PRO score.

Table 2. Parameters required by `%TTD`.

Parameter	Description
`dataset`	SAS dataset
`npat`	Patient identification (numeric or character format)
`time`	Time identification (numeric format), the baseline measure must be equal to 0
`date`	Date of patient-reported outcome (PRO) measurement (numeric or date format)
`score`	List of PRO score(s) (numeric format) (if several scores are treated simultaneously, use the « `|` » as a separator)
`MID`	List of threshold(s) for the minimal important difference (MID) (if several MID are considered, use the « `|` » as a separator)
`order`	List of value(s) of the `order` option (if several order are considered, use the « `|` » as a separator) `= 1 / = 2`
`ref`	`= 0 (by default) / = 1 / = 2`
`definition`	`= 1 (by default) / = 2`
`no_baseline`	`= 0 (by default) / = 1 / = 2`
`no_follow`	`= 0 (by default) / = 1`
`death`	`= 0 (by default) / = 1`
`date_death`	`= NA (by default) / = date of death variable` (numeric or date format)
`out`	Output dataset

Table 3. Parameters required by `%TUDD`.

Parameter	Description
`dataset`	SAS dataset
`npat`	Patient identification (numeric or character format)
`time`	Time identification (numeric format), the baseline measure must be equal to 0
`date`	Date of patient-reported outcome (PRO) measurement (numeric or date format)
`score`	List of PRO score(s) (numeric format) (if several scores are treated simultaneously, use the « `|` » as a separator)
`MID`	List of threshold(s) for the minimal important difference (MID) (if several MID are considered, use the « `|` » as a separator)
`order`	List of value(s) of the `order` option (if several order are considered, use the « `|` » as a separator) `= 1 / = 2`
`ref`	`= 0 (by default) / = 1 / = 2`
`definitive`	`= 1 (by default) / = 2 / = 3 / = 4`
`no_baseline`	`= 0 (by default) / = 1 / = 2`
`no_follow`	`= 0 (by default) / = 1`
`death`	`= 0 (by default) / = 1`
`date_death`	`= NA (by default) / = date of death variable` (numeric or date format)
`out`	Output dataset

Table 4. Output dataset (called « TTD_definition2 ») from `%TTD` macro.

Id	event_PF	time_PF	event_EF	time_EF	event_FA	time_FA	event_PA	time_PA	event_AP	time_AP
3	0	0.0328542094	0	0.0328542094	0	0.0328542094	0	0.0328542094	0	0.0328542094
4	.	.	.	.	.	.	.	.	.	.
8	.	.	.	.	.	.	.	.	.	.
9	0	2.3983572895	0	2.3983572895	0	2.3983572895	0	2.3983572895	0	2.3983572895
12	0	10.546201232	1	4.9281314168	1	0.9527720739	0	10.546201232	1	4.9281314168
13	1	2.7926078029	0	5.5852156057	0	5.5852156057	0	5.5852156057	0	5.5852156057
29	0	2.135523614	0	2.135523614	0	2.135523614	0	2.135523614	0	2.135523614
32	0	5.4866529774	0	5.4866529774	0	5.4866529774	1	0.9199178645	0	5.4866529774

PF, physical functioning; EF, emotional functioning; FA, fatigue; PA, pain; AP, appetite loss; A point represents a missing value. Time in months.

Table 5. Output dataset (called « TUDD_definitive2 ») from `%TUDD` macro.

Id	event_PF	time_PF	event_EF	time_EF	event_FA	time_FA	event_PA	time_PA	event_AP	time_AP
3	0	0.0328542094	0	0.0328542094	0	0.0328542094	0	0.0328542094	0	0.0328542094
4	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	1	1.2156057495	1	2.3983572895	1	2.3983572895	0	2.3983572895	1	2.3983572895
12	1	0.9527720739	0	10.546201232	0	10.546201232	0	10.546201232	0	10.546201232
13	1	5.5852156057	0	5.5852156057	0	5.5852156057	0	5.5852156057	0	5.5852156057
29	0	2.135523614	0	2.135523614	0	2.135523614	0	2.135523614	0	2.135523614
32	1	4.5667351129	0	5.4866529774	1	0.9199178645	0	5.4866529774	1	4.5667351129

PF, physical functioning; EF, emotional functioning; FA, fatigue; PA, pain; AP, appetite loss.
Time in months.

Supplementary Table A. SAS code for the `%TTD` macro.

```
/*****************************************************************************
* %TTD SAS macro program
*
* A SAS macro program to estimate the time to deterioration for patient-reported
* outcomes (PROs) data.
*
* Event definition: First or transient deterioration.
*
* Version 1.0 (18-05-2020)
*
* Emilie CHARTON,
* Methodology and Quality of Life Unit in Oncology (INSERM UMR 1098)
* University Hospital of Besancon, France
* E-mail: echarton@chu-besancon.fr
*
* Amelie ANOTA,
* French National Platform Quality of Life and Cancer
* Methodology and Quality of Life Unit in Oncology (INSERM UMR 1098)
* University Hospital of Besancon, France
* E-mail: aanota@chu-besancon.fr
*****************************************************************************/

/*****************************************************************************
Syntax:
%TTD(dataset=, *Name of the dataset;
     npat=, *Name of the patient identification number variable in the dataset;
     time=, *Name of the time identification variable in the dataset.
             The baseline measure must be equal to 0;
     date=, *Name of the date of PRO measurement variable in the dataset;
     score=, *List of name(s) of the PRO score(s) variable(s) in the dataset to be
              studied. Multiple PRO scores can be treated for one application of
              TTD macro, using the "|" command as a separator;
     MID=, *List of threshold(s) for the minimal important difference (MID).
            If x PRO scores are considered in the score option, then x MID must be
            considered using the "|" command as a separator;
     order=, *List of value(s) of order option (1/2):
              - 1: a deterioration is observed when the score decreases
              - 2: a deterioration is observed if the score increases
              If x PRO scores are considered in the score option, then x order must
              be considered using the "|" command as a separator;
     ref=, *0/1/2:
            - 0: the reference score is the baseline score (by default)
            - 1: the reference score is the best previous score
            - 2: the reference score is the immediately preceding score;
     definition=, *1/2:
                   - 1: first deterioration (i.e. a first deterioration of at least
                        one MID unit as compared to the reference score)
                       (by default)
                   - 2: transient deterioration (i.e. after the post-baseline
                        deterioration at least one follow-up score should be higher
                        than or at least be at the baseline level);
     no_baseline=, *0/1/2:
                    - 0: patients with no baseline score are censored at baseline,
                         i.e. day 0 (by default)
                    - 1: patients with no baseline score are deteriorated at
                         baseline
                    - 2: patients with no baseline score are excluded of the
                         analysis;
     no_follow=, *0/1:
                  - 0: patients with no follow-up score are censored just after
                       baseline, i.e. day 1 (by default)
                  - 1: patients with no follow-up score are deteriorated just after
                       baseline;
     death=, *0/1:
              - 0: patients who died during the study with no deterioration before
                   are censored at the time of the last available PRO measurement
```

```
                  (by default)
              - 1: all-cause death as an event at date of death;
     date_death=, *NA/date of death variable in the dataset:
                   - NA if death is censored (by default)
                   - date of death in other case;
     out=); *Name of the output dataset;

Libraries are allowed in the dataset= and out= options.
Missing values must be represented by a ".".
*****************************************************************************/


%macro TTD(dataset=,
           npat=,
           time=,
           date=,
           score=,
           MID=,
           order=,
           ref=0,
           definition=1,
           no_baseline=0,
           no_follow=0,
           death=0,
           date_death=NA,out=) / PARMBUFF;

%if %length(&dataset.)=0 %then %do;
    %put %str(WARNING: THE NAME OF DATASET MUST BE ENTERED IN THE DATASET OPTION.);
    %return;
%end;

%if %length(&npat.)=0 %then %do;
    %put %str(WARNING: MISSING VALUE FOR THE NPAT OPTION.);
    %return;
%end;

%if %length(&time.)=0 %then %do;
    %put %str(WARNING: MISSING VALUE FOR THE TIME OPTION.);
    %return;
%end;

%if %length(&date.)=0 %then %do;
    %put %str(WARNING: MISSING VALUE FOR THE DATE OPTION.);
    %return;
%end;

%if %sysfunc(countw(&score.,|)) ne %sysfunc(countw(&MID.,|))
 or %sysfunc(countw(&score.,|)) ne %sysfunc(countw(&order.,|))
 or %sysfunc(countw(&order.,|)) ne %sysfunc(countw(&MID.,|))
%then %do;
    %put %str(WARNING: THE NUMBER OF ELEMENTS ENTERED IN THE SCORE, MID AND ORDER
OPTIONS MUST BE EQUAL);
    %return;
%end;

%if %sysevalf(&ref.<0,integer) or (%sysevalf(&ref.>0,integer) and
    %sysevalf(&ref.<1,integer)) or (%sysevalf(&ref.>1,integer) and
    %sysevalf(&ref.<2,integer)) or %sysevalf(&ref.>2,integer) %then %do;
    %put %str(WARNING: THE REF PARAMETER MUST BE EQUAL TO 0, 1 or 2.);
    %return;
%end;

%if %sysevalf(&definition.<1,integer) or (%sysevalf(&definition.>1,integer) and
    %sysevalf(&definition.<2,integer)) or %sysevalf(&definition.>2,integer)
    %then %do;
    %put %str(WARNING: THE DEFINITION PARAMETER MUST BE EQUAL TO 1 OR 2.);
    %return;
%end;
```

```
%if %sysevalf(&no_baseline.<0,integer) or (%sysevalf(&no_baseline.>0,integer) and
    %sysevalf(&no_baseline.<1,integer)) or (%sysevalf(&no_baseline.>1,integer) and
    %sysevalf(&no_baseline.<2,integer)) or %sysevalf(&no_baseline.>2,integer)
    %then %do;
    %put %str(WARNING: THE NO_BASELINE PARAMETER MUST BE EQUAL TO 0, 1 OR 2.);
    %return;
%end;

%if %sysevalf(&no_follow.<0,integer) or (%sysevalf(&no_follow.>0,integer) and
    %sysevalf(&no_follow.<1,integer)) or %sysevalf(&no_follow.>1,integer)
    %then %do;
    %put %str(WARNING: THE NO_FOLLOW PARAMETER MUST BE EQUAL TO 0 OR 1.);
    %return;
%end;

%if %sysevalf(&death.<0,integer) or (%sysevalf(&death.>0,integer) and
    %sysevalf(&death.<1,integer)) or %sysevalf(&death.>1,integer) %then %do;
    %put %str(WARNING: THE DEATH PARAMETER MUST BE EQUAL TO 0 OR 1.);
    %return;
%end;

%if &death.=1 and &date_death.=NA or &death.=0 and &date_death. ne NA %then %do;
    %put %str(WARNING: THE PARAMETER ENTERED IN DEATH AND DATE_DEATH OPTIONS MUST
BE CONSISTENT.);
    %return;
%end;


%local i;

%do i=1 %to %sysfunc(countw(&score.,|));
    %let final_score=%sysfunc(strip(%qscan(&score,&i,|)));
    %let final_MID=%sysfunc(strip(%qscan(&MID,&i,|)));
    %let final_order=%sysfunc(strip(%qscan(&order,&i,|)));

    %if %sysevalf(&final_MID.<0.1,boolean) %then %do;
        %let number_score=%sysfunc(countw(&score,|));

        proc datasets nodetails nolist;
            delete &out. out_1-out_&number_score.;;
        quit;

        %put %str(WARNING: THE MID PARAMETER MUST BE A POSITIVE VALUE.);
        %return;
    %end;

    %if %sysevalf(&final_order.<1,integer) or (%sysevalf(&final_order.>1,integer)
        and
        %sysevalf(&final_order.<2,integer)) or %sysevalf(&final_order.>2,integer)
        %then %do;
        %let number_score=%sysfunc(countw(&score,|));

        proc datasets nodetails nolist;
            delete &out. out_1-out_&number_score.;;
        quit;

        %put %str(WARNING: THE ORDER PARAMETER MUST BE EQUAL TO 1 OR 2.);
        %return;
    %end;

    %ComputeTTD(_dataset=&dataset.,
                _npat=&npat.,
                _time=&time.,
                _date=&date.,
                _score=&final_score.,
                _MID=&final_MID.,
                _order=&final_order.,
```

```
                _ref=&ref.,
                _definition=&definition.,
                _no_baseline=&no_baseline.,
                _no_follow=&no_follow.,
                _death=&death.,
                _date_death=&date_death.,
                _out=&out.);

    data out_&i.;
        set &out.;
        rename event=event_&final_score. time=time_&final_score.;
    run;
%end;


%do;
    %let number_score=%sysfunc(countw(&score,|));

    data &out.;
        merge out_1-out_&number_score.;
        by &npat.;
    run;

    proc datasets nodetails nolist;
        delete out_1-out_&number_score.;
    quit;
%end;

%mend TTD;



%macro ComputeTTD(_dataset=,
                  _npat=,
                  _time=,
                  _date=,
                  _score=,
                  _MID=,
                  _order=,
                  _ref=0,
                  _definition=1,
                  _no_baseline=0,
                  _no_follow=0,
                  _death=0,
                  _date_death=NA,
                  _out=) / PARMBUFF;

data qlq;
    set &_dataset.;
    if missing(&_date.) then delete;
    if missing(&_score.) then delete;
run;


data qlq11 (keep=&_npat. &_score.0 &_date.0);
    set qlq;
    where &_time.=0;
    rename &_date.=&_date.0 &_score.=&_score.0;
run;


data qlq12;
    set qlq;
    where &_time. ne 0;
run;

proc transpose data=qlq12 out=qlq1 prefix=&_date.;
    by &_npat.;
```

```
    var &_date.;
run;

proc transpose data=qlq12 out=qlq2 prefix=&_score.;
    by &_npat.;
    var &_score.;
run;


data qlq3;
    merge qlq11 qlq2 (drop=_name_) qlq1 (drop=_name_);
    by &_npat.;
run;


data &_out. (keep=&_npat. event time);
    set qlq3;
    event=.;
    time=.;
    date_fin=max(of &_date.:);
    array vect{*} &_score.:;
    array vect2{*} &_date.:;

    if missing(vect(1)) then do;
        if &_no_baseline.=0 then event=999; else event=1;
        time=0;
    end;

    if nmiss(of &_score.:)=dim(vect)-1 and vect(1) ne . then do;
        if &_no_follow.=0 then event=0; else event=1;
        time=1;
    end;

    if event=. then do;
        do i=1 to dim(vect);
            if &_definition.=1 then do;
                if &_order.=1 and vect(1)-vect(i)>=&_MID. and vect(i) ne . and
                   event=. then do;
                    event=1;
                    time=vect2(i)-vect2(1);
                end;

                if &_ref.=1 and &_order.=1 and vect(1)-vect(i)<0 and vect(i) ne .
                then vect(1)=vect(i);


                if &_order.=2 and vect(1)-vect(i)<=-&_MID. and vect(i) ne . and
                   event=. then do;
                    event=1;
                    time=vect2(i)-vect2(1);
                end;

                if &_ref.=1 and &_order.=2 and vect(1)-vect(i)>0 and vect(i) ne .
                then vect(1)=vect(i);


                if &_ref.=2 and vect(i) ne . then vect(1)=vect(i);
            end;

            if &_definition.=2 then do;
                if &_order.=1 and vect(1)-vect(i)>=&_MID. and vect(i) ne . and
                   event=. then do;
                    transient='F';

                    do j=i+1 to dim(vect);
                        if vect(1)<=vect(j) and vect(j) ne . then transient='T';
                    end;
```

```
                    if transient='T' then do;
                        event=1;
                        time=vect2(i)-vect2(1);
                    end;
                end;

                if &_ref.=1 and &_order.=1 and vect(1)-vect(i)<0 and vect(i) ne .
                then vect(1)=vect(i);


                if &_order.=2 and vect(1)-vect(i)<=-&_MID. and vect(i) ne . and
                   event=. then do;
                    transient='F';

                    do j=i+1 to dim(vect);
                        if vect(1)>=vect(j) and vect(j) ne . then transient='T';
                    end;

                    if transient='T' then do;
                        event=1;
                        time=vect2(i)-vect2(1);
                    end;
                end;

                if &_ref.=1 and &_order.=2 and vect(1)-vect(i)>0 and vect(i) ne .
                then vect(1)=vect(i);


                if &_ref.=2 and vect(i) ne . then vect(1)=vect(i);
            end;
        end;
    end;

    if event=. then do;
        event=0;
        time=date_fin-vect2(1);
    end;

    if missing(vect(1)) and &_no_baseline.=2 then do;
        event=.;
        time=.;
    end;
run;


%if &_death.=1 %then %do;
    data &_out.;
        merge &_out. qlq (keep=&_npat. &_date. &_date_death.);
        by &_npat.;
        if first.&_npat.;
    run;

    data &_out. (keep=&_npat. event time);
        set &_out.;
        if event=0 and &_date_death. ne . then do;
            event=1;
            time=&_date_death.-&_date.;
        end;

        if event=1 and time=1 then do;
            if &_date_death. ne . then time=&_date_death.-&_date.;
        end;
    run;
%end;


data &_out.;
    set &_out.;
```

```
    if event=999 then event=0;
run;


data &_out.;
    set &_out.;
    time=time/30.4375;
run;


data qlq4 (keep=&_npat.);
    set &_dataset.;
    if first.&_npat. then output;
    by &_npat.;
run;

data &_out.;
    merge qlq4 &_out.;
    by &_npat.;
run;

data &_out.;
    set &_out.;
    if &_no_baseline.=0 and event=. then do;
        event=0;
        time=0;
    end;

    if &_no_baseline.=1 and event=. then do;
        event=1;
        time=0;
    end;
run;


proc datasets nodetails nolist;
    delete qlq qlq1 qlq11 qlq12 qlq2 qlq3 qlq4;
quit;

%mend ComputeTTD;


/*****************************************************************************
The output dataset contains:
- npat: patients identifier number,
- event: a dummy variable equals to 1 if the patient is deteriorated and 0 if not
- time: a time variable equals to the time to deterioration or the time to censure

If multiple scores are treated for one application of TTD macro then the variables
event and time are created by score.
*****************************************************************************/
```

Supplementary Table B. SAS code for the `%TUDD` macro.

```
/*****************************************************************************
* %TUDD SAS macro program
*
* A SAS macro program to estimate the time to deterioration for patient-reported
* outcomes (PROs) data.
*
* Event definition: Definitive deterioration.
*
* Version 1.0 (18-05-2020)
*
* Emilie CHARTON,
* Methodology and Quality of Life Unit in Oncology (INSERM UMR 1098)
* University Hospital of Besancon, France
* E-mail: echarton@chu-besancon.fr
*
* Amelie ANOTA,
* French National Platform Quality of Life and Cancer
* Methodology and Quality of Life Unit in Oncology (INSERM UMR 1098)
* University Hospital of Besancon, France
* E-mail: aanota@chu-besancon.fr
*****************************************************************************/

/*****************************************************************************
Syntax:
%TUDD(dataset=, *Name of the dataset;
      npat=, *Name of the patient identification number variable in the dataset;
      time=, *Name of the time identification variable in the dataset.
              The baseline measure must be equal to 0;
      date=, *Name of the date of PRO measurement variable in the dataset;
      score=, *List of name(s) of the PRO score(s) variable(s) in the dataset to be
               studied. Multiple PRO scores can be treated for one application of
               TUDD macro, using the "|" command as a separator;
      MID=, *List of threshold(s) for the minimal important difference (MID).
             If x PRO scores are considered in the score option, then x MID must be
             considered using the "|" command as a separator;
      order=, *List of value(s) of order option (1/2):
               - 1: a deterioration is observed when the score decreases
               - 2: a deterioration is observed if the score increases
               If x PRO scores are considered in the score option, then x order
               must be considered using the "|" command as a separator;
      ref=, *0/1/2:
             - 0: the reference score is the baseline score (by default)
             - 1: the reference score is the best previous score
             - 2: the reference score is the immediately preceding score;
      definitive=, *1/2/3/4:
                    - 1: the deterioration is definitive if there is no further
                         improvement of more than one MID unit as compared to the
                         reference score (by default)
                    - 2: the deterioration is definitive if thereafter maintaining
                         this deterioration of at least one MID unit for all
                         following scores (i.e. the deterioration is observed for
                         all the scores following the first deterioration)
                    - 3: the deterioration is definitive if there is no further
                         improvement of more than one MID unit as compared to the
                         score qualifying the deterioration (i.e. the score at the
                         time of the first deterioration observed)
                    - 4: the deterioration is definitive if there is no further
                         improvement of more than the score qualifying the
                         deterioration;
      no_baseline=, *0/1/2:
                     - 0: patients with no baseline score are censored at baseline,
                          i.e. day 0 (by default)
                     - 1: patients with no baseline score are deteriorated at
                          baseline
                     - 2: patients with no baseline score are excluded of the
                          analysis;
```

```
      no_follow=, *0/1:
                   - 0: patients with no follow-up score are censored just after
                        baseline, i.e. day 1 (by default)
                   - 1: patients with no follow-up score are deteriorated just
                        after baseline;
      death=, *0/1:
               - 0: patients who died during the study with no deterioration before
                    are censored at the time of the last available PRO measurement
                   (by default)
               - 1: all-cause death as an event at date of death;
      date_death=, *NA/date of death variable in the dataset:
                    - NA if death is censored (by default)
                    - date of death in other case;
      out=); *Name of the output dataset;

Libraries are allowed in the dataset= and out= options.
Missing values must be represented by a ".".
*************************************************************************/


%macro TUDD(dataset=,
            npat=,
            time=,
            date=,
            score=,
            MID=,
            order=,
            ref=0,
            definitive=1,
            no_baseline=0,
            no_follow=0,
            death=0,
            date_death=NA,
            out=) / PARMBUFF;

%if %length(&dataset.)=0 %then %do;
    %put %str(WARNING: THE NAME OF DATASET MUST BE ENTERED IN THE DATASET OPTION.);
    %return;
%end;

%if %length(&npat.)=0 %then %do;
    %put %str(WARNING: MISSING VALUE FOR THE NPAT OPTION.);
    %return;
%end;

%if %length(&time.)=0 %then %do;
    %put %str(WARNING: MISSING VALUE FOR THE TIME OPTION.);
    %return;
%end;

%if %length(&date.)=0 %then %do;
    %put %str(WARNING: MISSING VALUE FOR THE DATE OPTION.);
    %return;
%end;

%if %sysfunc(countw(&score.,|)) ne %sysfunc(countw(&MID.,|))
 or %sysfunc(countw(&score.,|)) ne %sysfunc(countw(&order.,|))
 or %sysfunc(countw(&order.,|)) ne %sysfunc(countw(&MID.,|))
%then %do;
    %put %str(WARNING: THE NUMBER OF ELEMENTS ENTERED IN THE SCORE, MID AND ORDER
OPTIONS MUST BE EQUAL);
    %return;
%end;

%if %sysevalf(&ref.<0,integer) or (%sysevalf(&ref.>0,integer) and
    %sysevalf(&ref.<1,integer)) or (%sysevalf(&ref.>1,integer) and
    %sysevalf(&ref.<2,integer)) or %sysevalf(&ref.>2,integer) %then %do;
    %put %str(WARNING: THE REF PARAMETER MUST BE EQUAL TO 0, 1 or 2.);
```

```
    %return;
%end;

%if %sysevalf(&definitive.<1,integer) or (%sysevalf(&definitive.>1,integer) and
    %sysevalf(&definitive.<2,integer)) or (%sysevalf(&definitive.>2,integer) and
    %sysevalf(&definitive.<3,integer)) or (%sysevalf(&definitive.>3,integer) and
    %sysevalf(&definitive.<4,integer)) or %sysevalf(&definitive.>4,integer) %then
%do;
    %put %str(WARNING: THE DEFINITIVE PARAMETER MUST BE EQUAL TO 1, 2, 3 or 4.);
    %return;
%end;

%if %sysevalf(&no_baseline.<0,integer) or (%sysevalf(&no_baseline.>0,integer) and
    %sysevalf(&no_baseline.<1,integer)) or (%sysevalf(&no_baseline.>1,integer) and
    %sysevalf(&no_baseline.<2,integer)) or %sysevalf(&no_baseline.>2,integer)
    %then %do;
    %put %str(WARNING: THE NO_BASELINE PARAMETER MUST BE EQUAL TO 0, 1 OR 2.);
    %return;
%end;

%if %sysevalf(&no_follow.<0,integer) or (%sysevalf(&no_follow.>0,integer) and
    %sysevalf(&no_follow.<1,integer)) or %sysevalf(&no_follow.>1,integer)
    %then %do;
    %put %str(WARNING: THE NO_FOLLOW PARAMETER MUST BE EQUAL TO 0 OR 1.);
    %return;
%end;

%if %sysevalf(&death.<0,integer) or (%sysevalf(&death.>0,integer) and
    %sysevalf(&death.<1,integer)) or %sysevalf(&death.>1,integer) %then %do;
    %put %str(WARNING: THE DEATH PARAMETER MUST BE EQUAL TO 0 OR 1.);
    %return;
%end;

%if &death.=1 and &date_death.=NA or &death.=0 and &date_death. ne NA %then %do;
    %put %str(WARNING: THE PARAMETER ENTERED IN DEATH AND DATE_DEATH OPTIONS MUST
BE CONSISTENT.);
    %return;
%end;


%local i;

%do i=1 %to %sysfunc(countw(&score.,|));
    %let final_score=%sysfunc(strip(%qscan(&score,&i,|)));
    %let final_MID=%sysfunc(strip(%qscan(&MID,&i,|)));
    %let final_order=%sysfunc(strip(%qscan(&order,&i,|)));

    %if %sysevalf(&final_MID.<0.1,boolean) %then %do;
        %let number_score=%sysfunc(countw(&score,|));

        proc datasets nodetails nolist;
            delete &out. out_1-out_&number_score.;;
        quit;

        %put %str(WARNING: THE MID PARAMETER MUST BE A POSITIVE VALUE.);
        %return;
    %end;

       %if %sysevalf(&final_order.<1,integer) or (%sysevalf(&final_order.>1,integer)
        and
        %sysevalf(&final_order.<2,integer)) or %sysevalf(&final_order.>2,integer)
        %then %do;
        %let number_score=%sysfunc(countw(&score,|));

        proc datasets nodetails nolist;
            delete &out. out_1-out_&number_score.;;
        quit;
```

```
        %put %str(WARNING: THE ORDER PARAMETER MUST BE EQUAL TO 1 OR 2.);
        %return;
    %end;

    %ComputeTUDD(_dataset=&dataset.,
                 _npat=&npat.,
                 _time=&time.,
                 _date=&date.,
                 _score=&final_score.,
                 _MID=&final_MID.,
                 _order=&final_order.,
                 _ref=&ref.,
                 _definitive=&definitive.,
                 _no_baseline=&no_baseline.,
                 _no_follow=&no_follow.,
                 _death=&death.,
                 _date_death=&date_death.,
                 _out=&out.);

    data out_&i.;
        set &out.;
        rename event=event_&final_score. time=time_&final_score.;
    run;
%end;


%do;
    %let number_score=%sysfunc(countw(&score,|));

    data &out.;
        merge out_1-out_&number_score.;
        by &npat.;
    run;

    proc datasets nodetails nolist;
        delete out_1-out_&number_score.;
    quit;
%end;

%mend TUDD;



%macro ComputeTUDD(_dataset=,
                   _npat=,
                   _time=,
                   _date=,
                   _score=,
                   _MID=,
                   _order=,
                   _ref=0,
                   _definitive=1,
                   _no_baseline=0,
                   _no_follow=0,
                   _death=0,
                   _date_death=NA,
                   _out=) / PARMBUFF;

data qlq;
    set &_dataset.;
    if missing(&_date.) then delete;
    if missing(&_score.) then delete;
run;


data qlq11 (keep=&_npat. &_score.0 &_date.0);
    set qlq;
    where &_time.=0;
```

```
    rename &_date.=&_date.0 &_score.=&_score.0;
run;


data qlq12;
    set qlq;
    where &_time. ne 0;
run;

proc transpose data=qlq12 out=qlq1 prefix=&_date.;
    by &_npat.;
    var &_date.;
run;

proc transpose data=qlq12 out=qlq2 prefix=&_score.;
    by &_npat.;
    var &_score.;
run;


data qlq3;
    merge qlq11 qlq2 (drop=_name_) qlq1 (drop=_name_);
    by &_npat.;
run;


data &_out. (keep=&_npat. event time);
    set qlq3;
    event=.;
    time=.;
    date_fin=max(of &_date.:);
    array vect{*} &_score.:;
    array vect2{*} &_date.:;

    if missing(vect(1)) then do;
        if &_no_baseline.=0 then event=999; else event=1;
        time=0;
    end;

    if nmiss(of &_score.:)=dim(vect)-1 and vect(1) ne . then do;
        if &_no_follow.=0 then event=0; else event=1;
        time=1;
    end;

    if event=. then do;
        do i=1 to dim(vect);
            if &_order.=1 and vect(1)-vect(i)>=&_MID. and vect(i) ne . and event=.
            then do;
                definition='T';

                do j=i+1 to dim(vect);
                    if &_definitive.=1 and vect(1)+&_MID.<vect(j) and vect(j) ne .
                       and
                       definition='T' then definition='F';
                    if &_definitive.=2 and vect(1)-&_MID.<vect(j) and vect(j) ne .
                       and
                       definition='T' then definition='F';
                    if &_definitive.=3 and vect(i)+&_MID.<vect(j) and vect(j) ne .
                       and
                       definition='T' then definition='F';
                    if &_definitive.=4 and vect(i)<vect(j) and vect(j) ne . and
                       definition='T' then definition='F';
                end;

                if definition='T' then do;
                    event=1;
                    time=vect2(i)-vect2(1);
                end;
```

```
            end;

            if &_ref.=1 and &_order.=1 and vect(1)-vect(i)<0 and vect(i) ne .
            then vect(1)=vect(i);


            if &_order.=2 and vect(1)-vect(i)<=-&_MID. and vect(i) ne . and
               event=. then do;
                definition='T';

                do j=i+1 to dim(vect);
                    if &_definitive.=1 and vect(1)-&_MID.>vect(j) and vect(j) ne .
                       and
                       definition='T' then definition='F';
                    if &_definitive.=2 and vect(1)+&_MID.>vect(j) and vect(j) ne .
                       and
                       definition='T' then definition='F';
                    if &_definitive.=3 and vect(i)-&_MID.>vect(j) and vect(j) ne .
                       and
                       definition='T' then definition='F';
                    if &_definitive.=4 and vect(i)>vect(j) and vect(j) ne . and
                       definition='T' then definition='F';
                end;

                if definition='T' then do;
                    event=1;
                    time=vect2(i)-vect2(1);
                end;
            end;

            if &_ref.=1 and &_order.=2 and vect(1)-vect(i)>0 and vect(i) ne .
            then vect(1)=vect(i);


            if &_ref.=2 and vect(i) ne . then vect(1)=vect(i);
        end;
    end;

    if event=. then do;
        event=0;
        time=date_fin-vect2(1);
    end;

    if missing(vect(1)) and &_no_baseline.=2 then do;
        event=.;
        time=.;
    end;
run;


%if &_death.=1 %then %do;
    data &_out.;
        merge &_out. qlq (keep=&_npat. &_date. &_date_death.);
        by &_npat.;
        if first.&_npat.;
    run;

    data &_out. (keep=&_npat. event time);
        set &_out.;
        if event=0 and &_date_death. ne . then do;
            event=1;
            time=&_date_death.-&_date.;
        end;

        if event=1 and time=1 then do;
            if &_date_death. ne . then time=&_date_death.-&_date.;
        end;
    run;
```

```
%end;


data &_out.;
    set &_out.;
    if event=999 then event=0;
run;


data &_out.;
    set &_out.;
    time=time/30.4375;
run;


data qlq4 (keep=&_npat.);
    set &_dataset.;
    if first.&_npat. then output;
    by &_npat.;
run;

data &_out.;
    merge qlq4 &_out.;
    by &_npat.;
run;

data &_out.;
    set &_out.;
    if &_no_baseline.=0 and event=. then do;
        event=0;
        time=0;
    end;

    if &_no_baseline.=1 and event=. then do;
        event=1;
        time=0;
    end;
run;


proc datasets nodetails nolist;
    delete qlq qlq1 qlq11 qlq12 qlq2 qlq3 qlq4;
quit;

%mend ComputeTUDD;


/*****************************************************************************
The output dataset contains:
- npat: patients identifier number,
- event: a dummy variable equals to 1 if the patient is deteriorated and 0 if not
- time: a time variable equals to the time to deterioration or the time to censure

If multiple scores are treated for one application of TUDD macro then the variables
event and time are created by score.
*****************************************************************************/
```

Supplementary Table C. SAS dataset (called « QoL_scores ») used to apply the macros.

Id	Time	DateQoL	PF	EF	FA	PA	AP	Arm
3	0	12/12/2013	100	58.333333333	66.666666667	100	0	1
4	2	14/01/2014	53.333333333	75	77.777777778	0	33.333333333	2
4	3	18/02/2014	53.333333333	66.666666667	55.555555556	0	0	2
4	4	04/03/2014	53.333333333	66.666666667	66.666666667	0	0	2
4	6	21/05/2014	46.666666667	66.666666667	44.444444444	0	0	2
8	.	.	.	.	.	.	.	1
9	0	06/01/2014	80	83.333333333	55.555555556	50	66.666666667	1
9	2	12/02/2014	73.333333333	83.333333333	55.555555556	16.666666667	66.666666667	1
9	3	20/03/2014	20	66.666666667	77.777777778	0	100	1
12	0	14/01/2014	86.666666667	66.666666667	33.333333333	16.666666667	66.666666667	2
12	2	12/02/2014	60	88.888888889	55.555555556	0	66.666666667	2
12	6	13/06/2014	73.333333333	58.333333333	33.333333333	0	100	2
12	7	15/07/2014	73.333333333	66.666666667	44.444444444	0	100	2
12	9	05/09/2014	80	83.333333333	33.333333333	0	66.666666667	2
12	10	06/10/2014	80	66.666666667	33.333333333	0	66.666666667	2
12	11	03/11/2014	80	66.666666667	33.333333333	0	33.333333333	2
12	12	01/12/2014	73.333333333	66.666666667	33.333333333	0	66.666666667	2
13	0	28/01/2014	60	83.333333333	55.555555556	33.333333333	100	1
13	4	23/04/2014	46.666666667	91.666666667	55.555555556	0	33.333333333	1
13	5	21/05/2014	66.666666667	83.333333333	33.333333333	0	33.333333333	1
13	7	17/07/2014	46.666666667	100	55.555555556	0	66.666666667	1
29	0	17/02/2014	20	33.333333333	100	100	100	2
29	3	23/04/2014	26.666666667	41.666666667	100	100	100	2
32	0	20/02/2014	100	58.333333333	22.222222222	0	0	1
32	1	20/03/2014	100	91.666666667	33.333333333	33.333333333	0	1
32	5	09/07/2014	80	83.333333333	44.444444444	33.333333333	33.333333333	1
32	6	06/08/2014	73.333333333	66.666666667	55.555555556	0	33.333333333	1

PF, physical functioning; EF, emotional functioning; FA, fatigue; PA, pain; AP, appetite loss; Arm 1, nab-paclitaxel plus gemcitabine treatment; Arm 2, nab-paclitaxel plus simplified leucovorin and fluorouracil treatment; A point represents a missing value.

3. Investigation de la méthode du temps jusqu'à détérioration

3.1. Cohorte de patientes atteintes d'un cancer du sein

<u>Article :</u> Time to health-related quality of life score deterioration at 1-year follow-up after types of immediate latissimus dorsi breast reconstructions: a prospective longitudinal cohort study in breast cancer patients.

Article soumis

Résumé

Objectif

A ce jour, l'impact sur la QdV au cours du temps, d'une reconstruction mammaire immédiate post-mastectomie, est peu documenté. L'objectif de cette étude était d'évaluer, dans une cohorte de patientes atteintes d'un cancer du sein, les effets de la reconstruction mammaire sur le temps jusqu'à détérioration d'un score de QdV jusqu'à un an après reconstruction par lambeau du muscle grand dorsal autologue (ALD) ou assisté par un implant (LDI).

Méthodes

La QdV a été évaluée par les questionnaires EORTC QLQ-C30 et QLQ-BR23 avant la chirurgie, à 3, 6 et 12 mois après l'opération. Le TJD a été défini comme l'intervalle de temps entre l'inclusion de la patiente et l'apparition d'une première détérioration cliniquement significative du score de QdV par rapport au score à l'inclusion. En raison de la non-comparabilité entre les deux groupes de reconstruction mammaire à l'inclusion, les analyses du TJD ont utilisé la méthode de la pondération par l'inverse du score de propension (« *inverse probability of treatment weighted* »), pour tenir compte des caractéristiques des patientes à l'inclusion associées au traitement reçu. Une analyse de sensibilité a été réalisée, utilisant le meilleur niveau de QdV déjà observé par la patiente (« *best previous score* ») comme score de référence, afin d'étudier l'impact potentiel d'un effet « *response shift* ». Des modèles de Cox univariés ont estimé les hazard ratios (HRs) et les intervalles de confiance (IC) à 99% du type de chirurgie.

Résultats

Entre 2005 et 2011, 230 patientes ont été recrutées dans cette cohorte : 105 patientes dans le groupe LDI et 125 patientes dans le groupe ALD. Le TJD était similaire pour presque toutes les dimensions explorées, à l'exception des dimensions perte d'appétit et constipation où les femmes subissant une reconstruction ALD se détériorent significativement plus tardivement que celles incluses dans le groupe LDI (HR = 0,54 [IC99% 0,31-0,95] et HR = 0,41 [IC99% 0,25-0,69], respectivement). Les résultats des analyses du TJD par type de chirurgie lors de l'analyse de sensibilité utilisant le « *best previous score* » comme score de référence étaient similaires à ceux obtenus utilisant le score à l'inclusion comme score de référence. L'analyse de sensibilité a montré que les résultats du TJD étaient significativement en faveur des femmes subissant une reconstruction ALD pour les dimensions perte d'appétit et constipation (HR = 0,56 [IC99% 0,33-0,96] et HR = 0,52 [IC99% 0,32-0,83], respectivement).

Conclusion

Les profils de QdV semblent être similaires entre les deux groupes de reconstruction mammaire sur les 12 premiers mois post-opératoires. Les résultats des analyses du TJD, utilisant le score à l'inclusion et le meilleur niveau de QdV déjà observé par la patiente comme score de référence, étaient semblables. Aucun effet « *response shift* » entre les deux types de chirurgie de reconstruction mammaire n'a pu être démontré. La méthodologie du temps jusqu'à détérioration d'un score de QdV est cliniquement importante et constitue une évaluation complémentaire pour les futures études de QdV utilisant de nouveaux questionnaires spécifiques à la reconstruction mammaire, tels que l'EORTC QLQ-BRECON23 et le BREAST-Q.

Original article

Time to health-related quality of life score deterioration at 1-year follow-up after types of immediate latissimus dorsi breast reconstructions: a prospective longitudinal cohort study in breast cancer patients

Emilie Charton[1,2], Amélie Anota[1,2,3], Franck Bonnetain[1,2,3], Zoe E. Winters[4].

[1]Methodology and Quality of Life Unit in Oncology, University Hospital of Besançon, Besançon, Besançon, France

[2]Univ. Bourgogne Franche-Comté, INSERM, EFS BFC, UMR1098, Interactions Hôte-Greffon-Tumeur/Ingénierie Cellulaire et Génique, France

[3]French National Platform Quality of Life and Cancer, Besançon, France

[4]Surgery and Interventional Trials Unit, Division of Surgery and Interventional Science, University College London, London, UK

Correspondence to: Ms Emilie Charton, Methodology and Quality of Life Unit in Oncology (INSERM UMR 1098), University Hospital of Besançon, 3 Boulevard Fleming, 25030 Besançon, France (e-mail: echarton@chu-besancon.fr)

ABSTRACT

Purpose: Evidence regarding the time-related impact of immediate breast reconstruction (BRR) on women's health-related quality of life (HRQoL) is lacking. Study aims were to evaluate the effects of BRR on the time to HRQoL score deterioration (TTD) over 1-year following types of latissimus dorsi (LD) BRR, either implant-assisted (LDI) or autologous (ALD) flaps.

Methods: HRQoL was assessed using the EORTC QLQ-C30 and QLQ-BR23 questionnaires before surgery, and at 3-, 6- and 12-months post-operatively. TTD was defined as the time between patient inclusion and the first clinically significant HRQoL score deterioration compared to the baseline score. Due to non-comparability between the two BRR groups at baseline, the TTD analyses used the inverse probability of treatment weighted method of the propensity score to account for the patients' baseline characteristics associated with received treatment. Univariate Cox models estimated hazard ratios (HRs) and 99% confidence intervals (CIs) of the type of surgery.

Results: Two hundred and thirty patients were recruited (105 LDI and 125 ALD). The TTD were similar for almost all the dimensions explored, except for appetite loss and constipation with significantly longer TTD for women undergoing ALD compared to the LDI group (HR=0.54 [99%CI 0.31-0.95] and HR=0.41 [99%CI 0.25-0.69], respectively).

Conclusion: The HRQoL profiles seem to be similar between the two BRR groups over 12 months. The TTD methodology is a clinically meaningful and a complementary evaluation in future HRQoL studies using newly developed BRR-specific HRQoL questionnaires, such as the EORTC QLQ-BRECON23 and BREAST-Q.

Keywords: Prospective cohort; Breast reconstructions; Health-related quality of life; Time to deterioration.

DECLARATIONS

Conflict of Interest: The authors declare that they have no conflict of interest.

INTRODUCTION

Breast cancer is the most common cancer in women with recommendations for mastectomy in about 40%. The loss of a breast can adversely affect women's health-related quality of life (HRQoL). The option of immediate breast reconstruction using various techniques is offered to all suitable women to ameliorate this negative aspect [1]. The National Institute for Health and Clinical Excellence recommends immediate breast reconstruction to all women undergoing a mastectomy [2]. Breast reconstruction aims to restore breast symmetry and body image [3].

Randomized controlled trials are considered to be optimal study designs on which to base treatment recommendations by reducing inherent selection biases. Randomization ensures a homogeneous distribution between different treatment groups, accounting for characteristics that might influence treatment decisions [4,5]. Most studies in breast reconstruction are observational and the feasibility of randomization is challenging [1,6]. Moreover, recommended types of reconstruction often depend on patients' characteristics and/or preferences, including the surgeon's experience [4,5,1,7]. The propensity score method is used to minimize the bias of a non-randomized study design. The propensity score is the probability of receiving a treatment according to the observed patients' baseline characteristics [8]. Four methods based on the propensity score have been proposed: matching, stratification, co-variate adjustment, and inverse probability of treatment weighting [9]. In analyses of time-to-event, the inverse probability of treatment weighting method using the propensity score is recommended [10].

One way to measure what matters to patients is using patient-reported outcomes measures, which are self-reported by patients to assess the effects of disease or treatment (or both) on symptoms, functioning, and overall HRQoL from their perspective [11]. Importantly, methods of analyzing longitudinal HRQoL data should include time-to-event models, such as the time to HRQoL score deterioration (TTD) [12]. A number of definitions for the TTD have been proposed depending on the cancer setting, the reference score, the minimal important difference (MID), whether or not the notion of deterioration is a definitive concept, the consideration or not of death as an event, and finally

censoring rules. In the adjuvant setting where patients have a high chance of surviving their cancer, current recommendations are to use the observed time to the first clinically significant HRQoL deterioration [13]. Different reference scores have been proposed to qualify the deterioration in HRQoL such as: baseline score, best previous score and immediately preceding score. The latter two reference scores have been proposed to explore the potential impact of a response shift effect [14], reflecting the patients' adaptation process to the disease and treatment.

This study aimed to evaluate the impact on the TTD of HRQoL over 1-year after types of immediate breast reconstruction most frequently performed in the United Kingdom (UK) at the time. These procedures comprised pedicled latissimus dorsi (LD) flaps, either implant-assisted (LDI) or autologous (ALD).

METHODS

Study design

This prospective longitudinal multicentre cohort study was conducted (National Research Ethics Committee Wiltshire: 05/Q2008/14) in six UK centres (Bristol, Cambridge, Glasgow, Hull, Swindon and York). Patients were recruited between January 2005 and August 2011. This study is an update of the initial cohort [15]. All women with TNM stages 0-II invasive breast cancer, including ductal carcinoma *in situ* were eligible. Only women with an LDI or ALD flap reconstruction were included. Exclusions comprised women with a previous cancer or bilateral disease, including those undergoing a delayed breast reconstruction and have been previously reported elsewhere [15]. Eligible women signed a written informed consent.

Health-related quality of life assessment

HRQoL was assessed using the European Organisation for Research and Treatment of Cancer

(EORTC) QLQ-C30 cancer-specific questionnaire [16] and the EORTC QLQ-BR23 breast cancer module [17]. The Hospital Anxiety and Depression Scale (HADS) generic questionnaire was used to assess patients-reported anxiety and depression levels [18].

The QLQ-C30 comprises 30 items and permits assessments of a global health status, 5 functional scales (physical, role, emotional, cognitive and social functioning) and 9 symptomatic scales (fatigue, nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhoea and financial difficulties). The QLQ-BR23 module includes 23 items that evaluates 4 functional scales (body image, sexual functioning, sexual enjoyment and future perspective) and 4 symptomatic scales (systematic therapy side effects, breast symptoms, arm symptoms and upset by hair loss). Scores were generated according to the EORTC scoring manual [19]. These scores vary from 0 (worst) to 100 (best) for the global health status and the functional scales, and from 0 (best) to 100 (worst) for the symptomatic scales.

The HADS questionnaire includes 14 items, comprising 7 items relating to anxiety and another 7 items for depression. One score is generated for each dimension. Both scores vary from 0 to 21. A score ranging from 0-7 represents normal levels of depression or anxiety, between 8-10 for moderate levels of depression or anxiety, and between 11-21 for severe levels of depression or anxiety.

Patients' characteristics

Explanatory variables used in the study included age at surgery, body mass index, educational level, employment situation, other significant medical illnesses (rheumatoid arthritis, clotting disorders or connective tissue diseases), presence of previous breast or axillary surgery, reason for surgery (screen-detected or symptomatic), type of axillary surgery (axillary lymph node dissections or sentinel node assisted sample), tumour characteristics and treatment received (including surgery, chemotherapy (neoadjuvant or adjuvant) and chest wall radiotherapy) as previously reported [15].

Follow-up

Women completed the questionnaires before surgery, and at 3-, 6- and 12-months after surgery. Early surgical complications were graded according to the Dindo-Clavien classification [20] and assessed up to 3 months after surgery.

Statistical analysis

Descriptive analysis at baseline

Quantitative variables were described by median and range. Qualitative variables were summarized by their absolute and relative frequencies.

The profile of missing data at baseline was explored by comparisons between patients with at least one available HRQoL score at baseline compared to those patients without any available HRQoL scores. The comparison was performed according to baseline clinical and socio-demographic variables, in order to determine whether baseline missing data were dependent on the characteristics of the patients, using the t-test or Mann-Whitney nonparametric test for quantitative variables, and χ^2 or Fisher's exact test for qualitative variables. P-values < 0.01 were considered as statistically significant.

Propensity score

Univariate logistic regression models were performed to determine which variables were associated with the treatment received. Using the Akaike Information Criterion, the area under the Receiver Operating Characteristic (ROC) curve and the Hosmer and Lemeshow Goodness-of-Fit test [21]: a multivariate model was selected to estimate a propensity score. An area under the ROC curve close to 1 (minimum recommended level higher than 0.7) indicates that the co-variates predict perfectly the assigned surgery. This model included all non-collinear variables and was described by the Odds Ratios and 99% confidence intervals (CIs). All the longitudinal analyses incorporated the inverse probability of treatment weighting method of the propensity score by assigning a weight to each

patient. Women undergoing ALD procedures received a weight equal to 1/propensity score and the women receiving an LDI reconstruction were given a weight equal to 1/(1-propensity score) [10].

Longitudinal analysis

Patients with at least one available baseline HRQoL score were included in the longitudinal analysis, corresponding to a modified intention-to-treat population [22]. The TTD was defined as the time between patient inclusion into the study and the first HRQoL score deterioration of at least one MID unit, compared to the baseline score [12]. Specific thresholds were used for the MID according to each HRQoL dimension: 1-point (cognitive functioning); 2-points (insomnia, appetite loss and financial difficulties); 3-points (emotional functioning and pain); 4-points (future perspective); 5-points (global health status, physical functioning, fatigue, nausea and vomiting, dyspnoea, constipation, diarrhoea, sexual functioning and sexual enjoyment); 6-points (social functioning); 7-points (role functioning); 9-points (arm symptoms); 12-points (breast symptoms); 13-points (body image); 15-points (systematic therapy side effects) [23-25]. Patients with one HRQoL assessment at baseline and without other follow-up measures were censored at a day after baseline. Patients with no deterioration before the end of the study were censored at the time of their last HRQoL assessment. A sensitivity analysis using the best previous score as the reference score was performed [13].

The estimation of the TTD was calculated using the Kaplan-Meier estimation method. Univariate Cox analyses were carried out to estimate the Hazard Ratios (HRs) and 99% CIs of the treatment effect. All baseline variables were tested by univariate Cox analysis. Quantitative variables were transformed into qualitative variables using the median values. The impact of the baseline levels of anxiety and depression on the TTD were also studied. The durations of time to: chemotherapy, radiotherapy and early surgical complications (up to 3 months) were tested in univariate Cox analyses as time-dependent variables.

Multivariate Cox analyses explored the independent factors influencing the TTD. Variables significant at a threshold of 10% by univariate analyses were eligible for the multivariate model. The type of

surgery was forced into the multivariate model. Identical multivariate models were constructed for the sensitivity analysis, i.e. using the best previous score as the reference score. Each HRQoL score was studied excluding 'upset by hair loss' due to the lack of sufficient patients and observed events.

All tests were two-sided and contingent on many statistical tests, thereby setting the statistical threshold at 1% to indicate statistical significance, in keeping with the previous publication [15].

All analyses were performed with SAS (version 9.3) (SAS Institute Inc., Cary, NC, USA) and R (version 3.3.1) software.

RESULTS

Study population

Among the 290 women assessed for eligibility, a total of 230 were recruited: 105 (45.7%) patients had LDI and 125 (54.3%) underwent ALD reconstructions (*Figure 1*). At baseline, 82 (78.1%) women undergoing LDI and 103 (82.4%) patients having ALD surgery completed HRQoL questionnaires. No differences in baseline HRQoL levels were observed between both types of breast reconstructions (*Supplementary Table S1*).

Missing data analysis

Baseline clinical and socio-demographic characteristics are summarized according to the availability or not of baseline HRQoL scores in *Table 1*. Patients with available baseline HRQoL data differed from those with missing baseline HRQoL scores in terms of the numbers of axillary lymph nodes removed (P=0.002). For all other variables, both groups presented similar characteristics.

Propensity score

Multivariate logistic regression analysis, showed the following independent factors: a high school educational level or less, being employed, presence of other medical illnesses, no previous breast or axillary surgery, a body mass index ≥ to 25.45 kg/m^2, an invasive breast cancer, and a moderate/severe level of depression were associated with ALD reconstruction, without reaching statistical significance (*Supplementary Table S2*). The area under the ROC curve was equal to 0.73 and the Hosmer and Lemeshow Goodness-of-Fit test was equal to 0.099, indicating the predictiveness of these co-variates for assigned surgery. This model was therefore retained to estimate the propensity score, with longitudinal analyses performed on these 172 patients.

Longitudinal analysis

In the longitudinal analyses using the baseline score as the reference score, TTD results were significantly and clinically in favor of women undergoing ALD breast reconstruction for appetite loss (HR=0.54 [99%CI 0.31-0.95]) and constipation (HR=0.41 [99%CI 0.25-0.69]). Similar trends were observed for 9 dimensions of QLQ-C30 and 6 dimensions of QLQ-BR23, but were not statistically significant in either surgery group. Conversely, the 5 remaining dimensions (namely, cognitive functioning, dyspnoea, insomnia, diarrhoea and sexual functioning) seemed to present a longer TTD for the LDI group, but without reaching significance (*Table 2* and *Table 3*).

Multivariate analyses showed significantly shorter TTD as compared to the baseline score among patients with the following (*Supplementary Table S3 and Supplementary Table S4*):

- A time to commencement of adjuvant chemotherapy for global health status (HR=1.93 [99%CI 1.26-2.95]), role functioning (HR=1.82 [99%CI 1.19-2.79]), cognitive functioning (HR=1.74 [99%CI 1.02-2.96]), social functioning (HR=2.31 [99%CI 1.51-3.55]), body image (HR=1.97 [99%CI 1.27-3.06]), sexual functioning (HR=1.94 [99%CI 1.16-3.25]), fatigue (HR=2.00 [99%CI 1.34-3.00]), nausea and vomiting (HR=3.21 [99%CI 1.89-5.43]), dyspnoea

(HR=2.11 [99%CI 1.18-3.79]) and systematic therapy side effects (HR=4.32 [99%CI 2.63-7.09]),

- a time to the onset of early surgical complications for role functioning (HR=1.67 [99%CI 1.14-2.45]), social functioning (HR=1.52 [99%CI 1.02-2.28]), body image (HR=1.84 [99%CI 1.23-2.73]), nausea and vomiting (HR=1.74 [99%CI 1.03-2.93]), dyspnoea (HR=3.93 [99%CI 2.19-7.06]), appetite loss (HR=2.31 [99%CI 1.27-4.20]), diarrhoea (HR=3.38 [99%CI 1.73-6.61]) and systematic therapy side effects (HR=1.70 [99%CI 1.06-2.71]).

Results also showed significantly longer TTD as compared to the baseline score among patients with the following:

- A sentinel node assisted sample for emotional functioning (HR=0.46 [99%CI 0.23-0.96]) and insomnia (HR=0.45 [99%CI 0.24-0.87]),
- included in ALD group for constipation (HR=0.41 [99%CI 0.24-0.70]).

Sensitivity analysis

In the sensitivity analyses using the best previous score as the reference score, TTD results were significantly in favor of women undergoing ALD breast reconstruction for appetite loss (HR=0.56 [99%CI 0.33-0.96]) and constipation (HR=0.52 [99%CI 0.32-0.83]). Similar trends were observed for 11 dimensions of QLQ-C30 and 5 dimensions of QLQ-BR23, but not statistically significant. Conversely, the 4 remaining dimensions (namely, cognitive functioning, pain, sexual functioning and sexual enjoyment) seemed to present a longer TTD for LDI group, without significance (*Table 2* and *Table 3*). Using the best previous score as the reference score, results of the TTD analysis by type of surgery were similar to those obtained using the baseline score as the reference score.

Multivariate analyses were in accordance with those obtained using the baseline score as the reference (*Supplementary Table S3 and Supplementary Table S4*). Results showed significantly longer TTD among patients with the following:

- A time to commencement of adjuvant chemotherapy for global health status (HR=1.79 [99%CI 1.20-2.67]), role functioning (HR=1.70 [99%CI 1.12-2.60]), social functioning

(HR=2.20 [99%CI 1.45-3.34]), body image (HR=2.01 [99%CI 1.30-3.10]), sexual functioning (HR=1.94 [99%CI 1.20-3.14]), fatigue (HR=1.76 [99%CI 1.18-2.61]), nausea and vomiting (HR=2.79 [99%CI 1.68-4.63]), dyspnoea (HR=2.13 [99%CI 1.23-3.71]) and systematic therapy side effects (HR=4.14 [99%CI 2.55-6.70]),

- a time to the onset of early surgical complications for role functioning (HR=1.71 [99%CI 1.17-2.48]), body image (HR=1.73 [99%CI 1.18-2.55]), nausea and vomiting (HR=1.94 [99%CI 1.17-3.19]), dyspnoea (HR=3.19 [99%CI 1.88-5.43]), appetite loss (HR=2.36 [99%CI 1.34-4.15]) and diarrhoea (HR=2.71 [99%CI 1.51-4.87]).

Results also showed significantly longer TTD as compared to the baseline score among patients included in the ALD group for social functioning (HR=0.63 [99%CI 0.42-0.94]) and constipation (HR=0.53 [99%CI 0.33-0.86]).

DISCUSSION

In this study, women undergoing ALD reported a longer TTD of appetite loss (HR=0.54 [99%CI 0.31-0.95] and HR=0.56 [99%CI 0.33-0.96], using baseline and best previous score as the reference scores, respectively) and constipation (HR=0.41 [99%CI 0.25-0.69] and HR=0.52 [99%CI 0.32-0.83], respectively), compared to implant-associated LD flap reconstructions, albeit most HRQoL dimensions were similar between either of the LD procedures. These findings persisted relative to longitudinal trajectories from either baseline or best previous reference scores for TTD using the cancer-specific (EORTC QLQ-C30) and breast cancer (EORTC QLQ-BR23) questionnaires. Neither of these questionnaires are surgery-specific, but as previously published [15], the results showed similar HRQoL between the two types of latissimus dorsi breast reconstruction for up to a year after surgery. The appetite loss and constipation dimensions seem to be more impacted by chemotherapy scheduling than by the type of breast reconstruction, but this hypothesis could not be demonstrated. In the prior publication, a generalized estimating equations model was used to analyze similar longitudinal HRQoL data, demonstrating that chemotherapy and early surgery complications were

significant independent predictors of worse HRQoL at 1-year post-reconstruction. In this study, re-analyses of the data using the TTD approach in HRQoL outcomes affirms the relevance of these adverse independent clinical contributors. Present findings are not directly comparable to previous publications, with some inter-study differences in patient populations. The described HRQoL methods using the TTD approach, enables complementary analyses of this prospective longitudinal cohort data. Indeed, the advantage of the TTD method over other longitudinal models is that it incorporates the MID, ensuring the clinical interpretability of the results [13].

The response shift effect is a change in the meaning of HRQoL over time and is particularly important in longitudinal studies. A response shift effect characterizes the process of patient adaptation in relation to their illness and its' treatment. Taking response shift effect into account produces a reliable and significant interpretation of MID for each HRQoL dimensions. We addressed this issue in the sensitivity analysis, using the best previous score as the reference score. Consistent results in TTD of HRQoL dimensions were shown based on both baseline and best previous scores. The latter has been proposed and is crucial in evaluating the potential effect of response shift [13], whose significance might usurp the use of the baseline score as the recommended reference score in future. The response shift effect based on the best previous score for HRQoL is analogous to RECIST criteria [26] for evaluating tumor response where the 'nadir' reflects treatment response. Baseline reference scores may be compromised by a number of factors supporting the evaluation of best previous scores for assessing the TTD and response shift effect. In fact, using this score as the reference for the patient, we considered that the reference HRQoL level for the patient was not necessarily her first HRQoL assessment just before surgery, but maybe the best level of HRQoL that she experienced before the current assessment. Since all HRQoL results were similar in TTD analyses based on either the best previous or baseline reference scores, it might be that there was no differential response shift effect between both types of breast reconstruction surgery. The described method aimed to potentially evaluate the response shift effect on the TTD [13]. These methods mandate a future simulation study to validate their applicability for assessing response shift associated with the TTD.

A limitation of these findings was the lack of a validated breast reconstruction-specific questionnaire

at study inception [1]. Recently, a breast reconstruction module, the EORTC QLQ-BRECON23 questionnaire has just been validated [27], and is optimally validated for use in parallel with the EORTC QLQ-C30 and the updated version of the QLQ-BR23, called the QLQ-BR45 [28]. The revised QLQ-BR45 was also not available at study inception. Both questionnaires used in this study are recommended as part of the breast cancer core outcome set proposed by the International Consortium for Health Outcomes Measurement [29]. The BREAST-Q is also validated in women undergoing mastectomy and breast reconstruction with normative data increasing clinical interpretation [30,29]. Future studies should aim to compare the EORTC QLQ-BRECON23 to the BREAST-Q in methodologically robust trials that are powered for effect size and numbers of events [31]. The longitudinal HRQoL methods herein will aim to evaluate the EORTC QLQ-BRECON23 in future studies on types of breast reconstruction to observe the impact on TTD of HRQoL dimensions over time.

The methods described herein are novel relative to existing papers on immediate breast reconstruction, albeit LD procedures are currently less commonly performed relative to implant-only and microvascular autologous flaps. Autologous breast reconstructions yield superior levels of satisfaction regarding breast cosmetic outcomes and lower risks of complications compared to implant-based procedures in women receiving adjuvant (postmastectomy) radiotherapy [32].

References

1. Winters, Z. E., Benson, J. R., & Pusic, A. L. (2010). A systematic review of the clinical evidence to guide treatment recommendations in breast reconstruction based on patient- reported outcome measures and health-related quality of life. *Annals of Surgery, 252*(6), 929-942, doi:10.1097/SLA.0b013e3181e623db.

2. Mayor, S. (2009). NICE updates guidance on early and advanced breast cancer. *BMJ (Clinical research ed.), 338*, b815.

3. Thiruchelvam, P. T. R., McNeill, F., Jallali, N., Harris, P., & Hogben, K. (2013). Post-mastectomy breast reconstruction. *BMJ (Clinical research ed.), 347*, f5903.

4. Dauplat, J., Kwiatkowski, F., Rouanet, P., Delay, E., Clough, K., Verhaeghe, J. L., et al. (2017). Quality of life after mastectomy with or without immediate breast reconstruction. *The British Journal of Surgery,* doi:10.1002/bjs.10537.

5. Lee, C., Sunu, C., & Pignone, M. (2009). Patient-reported outcomes of breast reconstruction after mastectomy: a systematic review. *Journal of the American College of Surgeons, 209*(1), 123-133, doi:10.1016/j.jamcollsurg.2009.02.061.

6. Winters, Z. E., Emson, M., Griffin, C., Mills, J., Hopwood, P., Bidad, N., et al. (2015). Learning from the QUEST multicentre feasibility randomization trials in breast reconstruction after mastectomy. *The British Journal of Surgery, 102*(1), 45-56, doi:10.1002/bjs.9690.

7. Winters, Z. E., Griffin, C., Horne, R., Bidad, N., & McCulloch, P. (2014). Barriers to accrue to clinical trials and possible solutions. *British Journal of Cancer, 111*(4), 637-639, doi:10.1038/bjc.2014.318.

8. Little, R. J., & Rubin, D. B. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annual Review of Public Health, 21*, 121-145, doi:10.1146/annurev.publhealth.21.1.121.

9. Austin, P. C. (2011). An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behavioral Research, 46*(3), 399-424, doi:10.1080/00273171.2011.568786.

10. Austin, P. C. (2013). The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in Medicine, 32*(16), 2837-2849, doi:10.1002/sim.5705.

11. Calvert, M., Kyte, D., Price, G., Valderas, J. M., & Hjollund, N. H. (2019). Maximising the impact of patient reported outcome assessment for patients and society. *BMJ (Clinical research ed.), 364*, k5267, doi:10.1136/bmj.k5267.

12. Hamidou, Z., Dabakuyo, T. S., Mercier, M., Fraisse, J., Causeret, S., Tixier, H., et al. (2011). Time to Deterioration in Quality of Life Score as a Modality of Longitudinal Analysis in Patients with Breast Cancer. *The Oncologist, 16*(10), 1458-1468, doi:10.1634/theoncologist.2011-0085.

13. Anota, A., Hamidou, Z., Paget-Bailly, S., Chibaudel, B., Bascoul-Mollevi, C., Auquier, P., et al. (2015). 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? *Quality of Life Research, 24*(1), 5-18.

14. Sprangers, M. A., & Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Social Science & Medicine (1982), 48*(11), 1507-1515.

15. Winters, Z. E., Haviland, J., Balta, V., Benson, J., Reece-Smith, A., Betambeau, N., et al. (2013). Integration of patient-reported outcome measures with key clinical outcomes after immediate latissimus dorsi breast reconstruction and adjuvant treatment. *The British Journal of Surgery, 100*(2), 240-251, doi:10.1002/bjs.8959.

16. Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., et al. (1993). The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *Journal of the National Cancer Institute, 85*(5), 365-376.

17. Sprangers, M. A., Groenvold, M., Arraras, J. I., Franklin, J., te Velde, A., Muller, M., et al. (1996). The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 14*(10), 2756-2768, doi:10.1200/JCO.1996.14.10.2756.

18. Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*(6), 361-370.

19. Fayers, P. M., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., Bottomley, A., et al. (2001). The EORTC QLQ-C30 Scoring Manual (3rd edition).

20. Dindo, D., Demartines, N., & Clavien, P.-A. (2004). Classification of surgical complications: a new proposal with evaluation in a cohort of 6336 patients and results of a survey. *Annals of Surgery, 240*(2), 205-213, doi:10.1097/01.sla.0000133083.54934.ae.

21. Lemeshow, S., & Hosmer, D. W., Jr. (1982). A review of goodness of fit statistics for use in the development of logistic regression models. *Am J Epidemiol, 115*(1), 92-106, doi:10.1093/oxfordjournals.aje.a113284.

22. Bonnetain, F., Fiteni, F., Efficace, F., & Anota, A. (2016). Statistical Challenges in the Analysis of Health-Related Quality of Life in Cancer Clinical Trials. *Journal of Clinical Oncology, 34*(16), 1953-1956, doi:10.1200/JCO.2014.56.7974.

23. Cocks, K., King, M. T., Velikova, G., de Castro, G., Martyn St-James, M., Fayers, P. M., et al. (2012). Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *European Journal of Cancer (Oxford, England: 1990), 48*(11), 1713-1721, doi:10.1016/j.ejca.2012.02.059.

24. Osoba, D., Rodrigues, G., Myles, J., Zee, B., & Pater, J. (1998). Interpreting the significance of changes in health-related quality-of-life scores. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 16*(1), 139-144.

25. Ousmen, A., Conroy, T., Guillemin, F., Velten, M., Jolly, D., Mercier, M., et al. (2016). Impact of the occurrence of a response shift on the determination of the minimal important difference in a health-related quality of life score over time. *Health and Quality of Life Outcomes, 14*(1), 167, doi:10.1186/s12955-016-0569-5.

26. Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., Ford, R., et al. (2009). New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *European Journal of Cancer, 45*(2), 228-247, doi:10.1016/j.ejca.2008.10.026.

27. Winters, Z. E., Afzal, M., Rutherford, C., Holzner, B., Rumpold, G., da Costa Vieira, R. A., et al. (2018). International validation of the European Organisation for Research and Treatment of Cancer QLQ-BRECON23 quality-of-life questionnaire for women undergoing breast reconstruction. *The British Journal of Surgery, 105*(3), 209-222, doi:10.1002/bjs.10656.

28. Bjelic-Radisic, V., Cardoso, F., Cameron, D., Brain, E., Kuljanic, K., da Costa, R. A., et al. (2020). An international update of the EORTC questionnaire for assessing quality of life in breast cancer patients: EORTC QLQ-BR45. *Ann Oncol, 31*(2), 283-288, doi:10.1016/j.annonc.2019.10.027.

29. Ong, W. L., Schouwenburg, M. G., van Bommel, A. C. M., Stowell, C., Allison, K. H., Benn, K. E., et al. (2017). A Standard Set of Value-Based Patient-Centered Outcomes for Breast Cancer: The International Consortium for Health Outcomes Measurement (ICHOM) Initiative. *JAMA Oncol, 3*(5), 677-685, doi:10.1001/jamaoncol.2016.4851.

30. Mundy, L. R., Homa, K., Klassen, A. F., Pusic, A. L., & Kerrigan, C. L. (2017). Normative Data for Interpreting the BREAST-Q: Augmentation. *Plast Reconstr Surg, 139*(4), 846-853, doi:10.1097/PRS.0000000000003186.

31. Winters, Z. E., Afzal, M., Balta, V., Freeman, J., Llewellyn-Bennett, R., Rayter, Z., et al. (2016). Patient-reported outcomes and their predictors at 2- and 3-year follow-up after immediate latissimus dorsi breast reconstruction and adjuvant treatment. *The British Journal of Surgery, 103*(5), 524-536, doi:10.1002/bjs.10102.

32. Jagsi, R., Momoh, A. O., Qi, J., Hamill, J. B., Billig, J., Kim, H. M., et al. (2018). Impact of Radiotherapy on Complications and Patient-Reported Outcomes After Breast Reconstruction. *J Natl Cancer Inst, 110*(2), doi:10.1093/jnci/djx148.

Figure 1. Progress of patients through the study. LDI, implant-assisted latissimus dorsi breast reconstruction; ALD, autologous latissimus dorsi breast reconstruction. †A patient died before 12-month completion questionnaires, 81 eligible patients responded in the LDI group at 12-month.

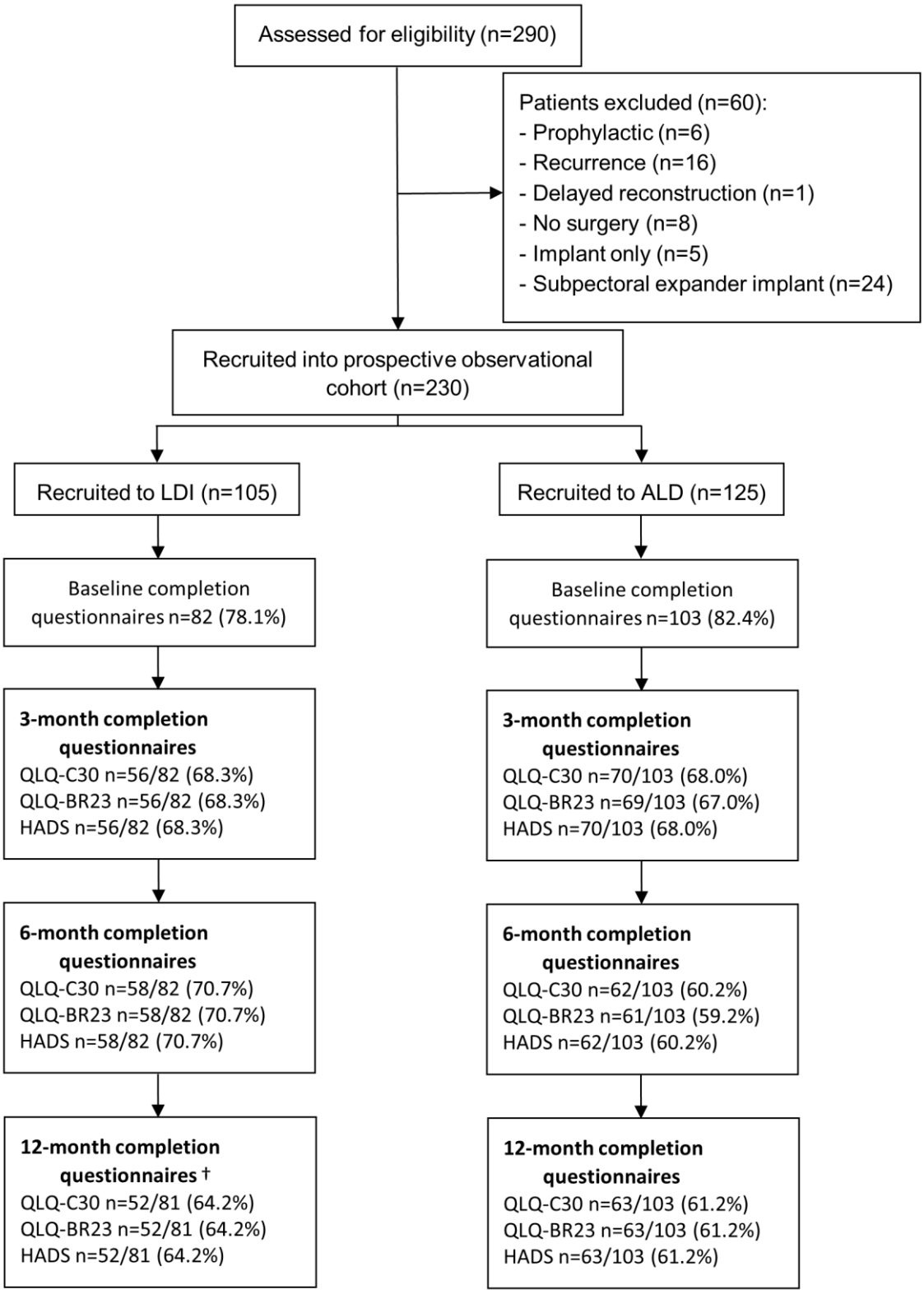


Table 1. Baseline clinical and socio-demographic characteristics of patients according to the availability of their baseline health-related quality of life questionnaire.

	Baseline HRQoL available (n=185)		HRQoL not available (n=45)		
	n	**%**	**n**	**%**	**P**
Type of surgery					
LDI	82	44.3	23	51.1	0.412
ALD	103	55.7	22	48.9	
Education					
≤ high school[1]	103	56.3	12	57.1	0.940
> high school[2]	80	43.7	9	42.9	
Employed					
Yes	143	78.1	21	100.0	0.017*
No	40	21.9	0	0.0	
Other medical illnesses[3]					
Yes	30	16.6	5	11.6	0.422
No	151	83.4	38	88.4	
Previous breast or axillary surgery					
Yes	73	39.9	10	23.3	0.042
No	110	60.1	33	76.7	
Neoadjuvant chemotherapy					
Yes	22	12.1	8	18.6	0.258
No	160	87.9	35	81.4	
Reason for surgery					
Screen-detected	81	44.5	15	34.1	0.210
Symptomatic	101	55.5	29	65.9	
Type of axillary surgery					
Axillary lymph nodes dissections	130	71.0	38	84.4	0.192*
Sentinel node assisted sample	42	23.0	5	11.1	
None	11	6.0	2	4.4	
Type of tumour					
Invasive only	36	19.8	6	13.3	0.263
DCIS only	45	24.7	8	17.8	
Invasive + DCIS	101	55.5	31	68.9	
Grade of tumour (invasive)					
Low/Intermediate	94	69.6	23	63.9	0.510
High	41	30.4	13	36.1	
Grade of tumour (DCIS)					
Low/Intermediate	43	34.1	11	31.4	0.765
High	83	65.9	24	68.6	
Lymph node positivity					
Yes	61	33.1	21	46.7	0.090
No	123	66.9	24	53.3	
Age at surgery (years)[4]	185	51.3 (22-74)	45	49.5 (31-71)	0.320**
Body mass index (kg/m^2)[4]	181	25.5 (17-41)	43	25.4 (19-38)	0.880***
Tumour size invasive (mm)[4]	134	18.0 (1-100)	35	21.0 (2-65)	0.158***
Tumour size DCIS (mm)[4]	124	30.0 (1-190)	30	26.5 (1-70)	0.615***
Number of axillary lymph nodes removed[4]	184	8.0 (0-31)	45	11.0 (0-36)	0.002***
Number of axillary lymph nodes involved[4]	184	0.0 (0-26)	45	0.0 (0-14)	0.104***

A χ^2 is used unless indicated otherwise; *Fisher's exact test; **t-test; ***Mann-Whitney non-parametric test. LDI, implant-assisted latissimus dorsi breast reconstruction; ALD, autologous latissimus dorsi breast reconstruction; DCIS, Ductal carcinoma *in* situ; [1]Including the secondary school and before it; [2]Including the Advanced level and after it; [3]Other systematic conditions such as rheumatoid arthritis, clotting disorders, connective tissue diseases; [4]median (range) for continuous variables.

Table 2. Results of the univariate Cox regression of the time to health-related quality of life score deterioration (TTD) for QLQ-C30 scores and comparison between the types of surgery.

	TTD as compared to the baseline score			TTD as compared to the best previous score		
	n (events)	Median (99%CI)	HR (99%CI)	n (events)	Median (99%CI)	HR (99%CI)
Global health status						
LDI	75 (42)	4.53 (3.88-6.80)	1	75 (51)	4.53 (3.88-6.77)	1
ALD	94 (50)	6.80 (4.30-13.10)	0.81 (0.55-1.17)	96 (57)	6.14 (4.30-11.83)	0.76 (0.53-1.07)
Physical functioning						
LDI	75 (46)	3.88 (3.61-6.34)	1	75 (48)	3.88 (3.61-6.08)	1
ALD	96 (65)	4.21 (3.84-6.54)	0.98 (0.70-1.38)	97 (68)	4.21 (3.88-6.47)	0.96 (0.69-1.35)
Role functioning						
LDI	75 (39)	4.53 (3.81-7.23)	1	75 (40)	4.53 (3.81-7.23)	1
ALD	96 (52)	5.98 (4.21-11.96)	0.93 (0.64-1.35)	97 (56)	5.98 (4.21-10.64)	0.97 (0.68-1.40)
Emotional functioning						
LDI	75 (27)	NA (6.80-NA)	1	75 (39)	6.90 (5.98-12.90)	1
ALD	93 (28)	14.80 (12.40-NA)	0.68 (0.42-1.10)	96 (44)	12.10 (11.83-12.80)	0.73 (0.49-1.08)
Cognitive functioning						
LDI	75 (24)	NA (12.60-NA)	1	75 (33)	12.40 (6.80-13.00)	1
ALD	93 (32)	13.10 (12.10-NA)	1.07 (0.67-1.71)	96 (42)	12.00 (6.67-NA)	1.06 (0.71-1.58)
Social functioning						
LDI	75 (39)	4.01 (3.61-NA)	1	75 (42)	4.01 (3.61-7.29)	1
ALD	93 (46)	7.62 (6.34-NA)	0.79 (0.53-1.17)	96 (49)	6.67 (6.34-NA)	0.73 (0.50-1.06)
Fatigue						
LDI	75 (45)	4.40 (3.81-6.60)	1	75 (50)	4.40 (3.81-6.51)	1
ALD	96 (57)	5.98 (4.21-7.62)	0.87 (0.61-1.24)	97 (66)	5.85 (4.21-6.67)	0.95 (0.68-1.33)
Nausea and vomiting						
LDI	75 (22)	NA (12.40-NA)	1	75 (24)	NA (6.87-NA)	1
ALD	96 (27)	NA (NA-NA)	0.65 (0.39-1.09)	97 (31)	NA (12.71-NA)	0.70 (0.43-1.15)
Pain						
LDI	75 (38)	6.90 (4.40-13.00)	1	75 (39)	6.90 (4.40-12.90)	1
ALD	96 (51)	7.52 (5.82-12.80)	0.96 (0.65-1.41)	97 (57)	6.97 (5.65-12.00)	1.06 (0.73-1.54)
Dyspnoea						
LDI	75 (20)	13.50 (13.50-NA)	1	75 (22)	13.50 (12.80-NA)	1
ALD	96 (30)	NA (12.80-NA)	1.03 (0.62-1.73)	97 (33)	12.80 (12.80-NA)	0.98 (0.60-1.60)
Insomnia						
LDI	75 (32)	12.70 (11.70-NA)	1	75 (38)	11.70 (6.44-12.70)	1
ALD	96 (38)	12.40 (11.80-13.20)	1.04 (0.68-1.59)	97 (42)	12.20 (7.46-13.20)	0.94 (0.63-1.41)
Appetite loss						
LDI	75 (20)	NA (12.40-NA)	1	75 (23)	NA (12.10-NA)	1
ALD	96 (19)	NA (NA-NA)	0.54 (0.31-0.95)	97 (22)	NA (NA-NA)	0.56 (0.33-0.96)
Constipation						
LDI	75 (29)	11.90 (6.80-NA)	1	75 (31)	11.80 (6.80-NA)	1
ALD	94 (22)	NA (NA-NA)	0.41 (0.25-0.69)	96 (25)	NA (12.80-NA)	0.52 (0.32-0.83)
Diarrhoea						
LDI	75 (14)	NA (NA-NA)	1	75 (20)	13.40 (13.20-NA)	1
ALD	93 (17)	NA (13.00-NA)	1.09 (0.59-2.01)	96 (19)	NA (13.00-NA)	0.85 (0.49-1.47)
Financial difficulties						
LDI	75 (19)	NA (NA-NA)	1	75 (19)	NA (NA-NA)	1
ALD	92 (21)	NA (NA-NA)	0.63 (0.36-1.13)	96 (24)	NA (13.10-NA)	0.75 (0.43-1.30)

LDI, implant-assisted latissimus dorsi breast reconstruction; ALD, autologous latissimus dorsi breast reconstruction; HR, hazard ratio; 99%CI, 99% confidence interval; NA, Not Available.

Table 3. Results of the univariate Cox regression of the time to health-related quality of life score deterioration (TTD) for QLQ-BR23 scores and comparison between the types of surgery.

	TTD as compared to the baseline score			TTD as compared to the best previous score		
	n (events)	**Median (99%CI)**	**HR (99%CI)**	**n (events)**	**Median (99%CI)**	**HR (99%CI)**
Body image						
LDI	72 (38)	6.74 (4.53-NA)	1	74 (40)	6.74 (4.53-6.90)	1
ALD	94 (50)	11.96 (6.18-12.60)	0.89 (0.60-1.31)	97 (53)	11.83 (6.18-12.60)	0.90 (0.62-1.32)
Sexual functioning						
LDI	74 (22)	NA (NA-NA)	1	75 (31)	13.00 (11.89-NA)	1
ALD	88 (35)	12.30 (6.41-NA)	1.57 (0.96-2.56)	93 (42)	11.60 (6.34-NA)	1.32 (0.86-2.03)
Sexual enjoyment						
LDI	50 (15)	NA (13.00-NA)	1	58 (16)	NA (13.00-NA)	1
ALD	48 (11)	13.10 (12.80-NA)	0.94 (0.47-1.90)	62 (14)	13.10 (12.80-NA)	1.04 (0.54-2.00)
Future perspective						
LDI	73 (18)	14.30 (NA-NA)	1	74 (32)	12.50 (11.80-NA)	1
ALD	94 (15)	NA (NA-NA)	0.64 (0.34-1.21)	97 (25)	NA (12.60-NA)	0.63 (0.39-1.03)
Systematic therapy side effects						
LDI	74 (27)	12.80 (6.80-NA)	1	75 (30)	12.60 (6.80-NA)	1
ALD	95 (33)	NA (12.20-NA)	0.87 (0.55-1.38)	97 (36)	14.10 (11.80-NA)	0.85 (0.55-1.32)
Breast symptoms						
LDI	72 (27)	NA (5.88-NA)	1	74 (28)	NA (6.44-NA)	1
ALD	93 (23)	NA (NA-NA)	0.62 (0.38-1.01)	97 (26)	NA (12.09-NA)	0.67 (0.41-1.07)
Arm symptoms						
LDI	72 (43)	4.53 (3.61-6.90)	1	74 (48)	4.53 (3.81-6.90)	1
ALD	93 (57)	6.34 (4.83-8.05)	0.94 (0.66-1.35)	97 (61)	6.34 (4.83-7.46)	0.92 (0.65-1.31)

LDI, implant-assisted latissimus dorsi breast reconstruction; ALD, autologous latissimus dorsi breast reconstruction; HR, hazard ratio; 99%CI, 99% confidence interval; NA, Not Available.

Supplementary Table S1. Description of health-related quality of life scores at baseline by type of surgery.

	LDI (n=82)			ALD (n=103)		
	n	**median**	**range**	**n**	**median**	**range**
QLQ-C30						
Global health status	82	75.0	17-100	100	75.0	42-100
Physical functioning	82	100.0	60-100	102	100.0	60-100
Role functioning	82	100.0	0-100	102	100.0	17-100
Emotional functioning	82	66.7	0-100	99	75.0	0-100
Cognitive functioning	82	83.3	0-100	99	83.3	0-100
Social functioning	82	83.3	0-100	98	100.0	0-100
Fatigue	82	22.2	0-100	102	22.2	0-89
Nausea and vomiting	82	0.0	0-33	102	0.0	0-50
Pain	82	0.0	0-50	102	0.0	0-83
Dyspnoea	82	0.0	0-100	102	0.0	0-67
Insomnia	82	33.3	0-100	102	33.3	0-100
Appetite loss	82	0.0	0-100	102	0.0	0-100
Constipation	82	0.0	0-100	100	0.0	0-100
Diarrhoea	82	0.0	0-100	99	0.0	0-100
Financial difficulties	82	0.0	0-100	97	0.0	0-100
QLQ-BR23						
Body image	79	83.3	0-100	99	91.7	17-100
Sexual functioning	81	33.3	0-100	93	33.3	0-83
Sexual enjoyment	53	66.7	33-100	51	66.7	0-100
Future perspective	80	33.3	0-100	99	66.7	0-100
Systematic therapy side effects	81	9.5	0-43	100	9.5	0-52
Breast symptoms	78	8.3	0-75	98	8.3	0-75
Arm symptoms	79	0.0	0-67	98	0.0	0-44
Upset by hair loss	7	33.3	0-67	9	66.7	0-100
HADS						
Normal level of anxiety	37	6.0	1-7	48	5.0	0-7
Moderate level of anxiety	17	9.0	8-10	28	9.0	8-10
Severe level of anxiety	27	15.0	11-21	25	14.0	11-21
Normal level of depression	73	2.0	0-7	88	1.0	0-7
Moderate level of depression	5	10.0	8-10	10	9.0	8-10
Severe level of depression	3	15.0	12-15	3	14.0	11-16

LDI, implant-assisted latissimus dorsi breast reconstruction; ALD, autologous latissimus dorsi breast reconstruction; QLQ-C30, Quality of Life Questionnaire-Core 30; QLQ-BR23, Quality of Life Questionnaire-Breast Cancer 23; HADS, Hospital Anxiety and Depression Scale.

Supplementary Table S2. Results of the multivariate logistic regression between implant-assisted latissimus dorsi and autologous latissimus dorsi breast reconstructions (n=172).

		OR	99%CI	P
Education	(> high school) vs. ≤ high school	0.50	0.21-1.21	0.043
Employed	yes vs. no	2.73	0.96-7.77	0.014
Other medical illnesses	yes vs. no	1.58	0.45-5.54	0.346
Previous breast or axillary surgery	yes vs. no	0.71	0.29-1.73	0.324
Body mass index (kg/m^2)	(≥ 25.45) vs. < 25.45	1.68	0.70-4.01	0.128
Type of tumour	DCIS vs. Invasive	0.28	0.07-1.06	0.014
	Invasive+DCIS vs. Invasive	0.72	0.22-2.30	0.464
Depression	Moderate/Severe vs. Normal	1.46	0.34-6.29	0.500

OR, odds ratio; 99%CI, 99% confidence interval; DCIS, Ductal carcinoma *in* situ; Education, ≤ high school includes the secondary school and before it, > high school includes the Advanced level and after it; Other medical illnesses, other systematic conditions such as rheumatoid arthritis, clotting disorders, connective tissue diseases.

Supplementary Table S3. Results of the multivariate Cox regression of the time to health-related quality of life score deterioration (TTD) for the global health status and functional scales of the QLQ-C30 and QLQ-BR23 questionnaires.

		TTD as compared to the baseline score				TTD as compared to the best previous score			
		n (events)	**HR**	**99%CI**	**P**	**n (events)**	**HR**	**99%CI**	**P**
Global health status		169 (92)				171 (108)			
Type of surgery	ALD vs. LDI		0.76	0.52-1.13	0.074		0.72	0.50-1.04	0.020
Time to chemotherapy	Time-dependent variable		1.93	1.26-2.95	< 0.001		1.79	1.20-2.67	< 0.001
Time to early surgical complications	Time-dependent variable		1.27	0.86-1.86	0.114		1.24	0.87-1.77	0.125
Physical functioning		171 (111)				172 (116)			
Type of surgery	ALD vs. LDI		1.04	0.74-1.48	0.751		1.04	0.73-1.46	0.795
Type of tumour					0.009				0.006
	DCIS vs. Invasive		1.20	0.69-2.07	0.402		1.24	0.73-2.13	0.297
	Invasive+DCIS vs. Invasive		1.67	1.02-2.74	0.008		1.72	1.06-2.80	0.004
Depression	Moderate/Severe vs. Normal		1.38	0.84-2.27	0.099		1.32	0.80-2.17	0.150
Time to early surgical complications	Time-dependent variable		1.27	0.89-1.80	0.086		1.31	0.93-1.86	0.045
Role functioning		171 (91)				172 (96)			
Type of surgery	ALD vs. LDI		0.94	0.63-1.40	0.675		0.99	0.67-1.46	0.949
Previous breast or axillary surgery	yes vs. no		0.67	0.45-1.01	0.011		0.63	0.42-0.94	0.003
Anxiety					0.009				0.012
	Moderate vs. Normal		1.52	0.95-2.45	0.023		1.53	0.96-2.42	0.018
	Severe vs. Normal		1.63	1.04-2.53	0.005		1.56	1.01-2.41	0.009
Time to chemotherapy	Time-dependent variable		1.82	1.19-2.79	< 0.001		1.70	1.12-2.60	0.001
Time to early surgical complications	Time-dependent variable		1.67	1.14-2.45	0.001		1.71	1.17-2.48	< 0.001
Emotional functioning		157 (51)				160 (79)			
Type of surgery	ALD vs. LDI		0.62	0.37-1.06	0.021		0.76	0.49-1.17	0.101
Education	(> high school) vs. ≤ high school		1.36	0.81-2.26	0.124		1.43	0.94-2.17	0.029
Type of axillary surgery	Sentinel node assisted sample vs. AxLND		0.46	0.23-0.96	0.006		0.70	0.42-1.14	0.060
Anxiety					0.037				0.853
	Moderate vs. Normal		1.01	0.56-1.83	0.973		1.10	0.65-1.86	0.629
	Severe vs. Normal		0.51	0.25-1.03	0.014		1.09	0.66-1.80	0.657
Time to early surgical complications	Time-dependent variable		1.41	0.82-2.41	0.100		1.43	0.93-2.22	0.034
Cognitive functioning		168 (56)				171 (75)			
Type of surgery	ALD vs. LDI		1.10	0.68-1.78	0.609		1.09	0.72-1.65	0.592
Previous breast or axillary surgery	yes vs. no		1.70	1.05-2.74	0.004		1.24	0.81-1.88	0.190
Depression	Moderate/Severe vs. Normal		0.23	0.05-0.97	0.009		0.98	0.51-1.90	0.945
Time to chemotherapy	Time-dependent variable		1.74	1.02-2.96	0.008		1.41	0.87-2.27	0.068
Time to early surgical complications	Time-dependent variable		1.48	0.91-2.38	0.037		1.43	0.95-2.16	0.023
Social functioning		168 (85)				171 (91)			
Type of surgery	ALD vs. LDI		0.69	0.46-1.05	0.022		0.63	0.42-0.94	0.003
Education	(> high school) vs. ≤ high school		1.19	0.80-1.77	0.259		1.15	0.78-1.68	0.361
Previous breast or axillary surgery	yes vs. no		0.66	0.43-1.01	0.013		0.60	0.39-0.91	0.001
Depression	Moderate/Severe vs. Normal		0.50	0.21-1.19	0.039		0.81	0.42-1.54	0.391
Time to chemotherapy	Time-dependent variable		2.31	1.51-3.55	< 0.001		2.20	1.45-3.34	< 0.001
Time to early surgical complications	Time-dependent variable		1.52	1.02-2.28	0.008		1.40	0.95-2.05	0.026
Body image		166 (88)				171 (93)			
Type of surgery	ALD vs. LDI		0.79	0.53-1.20	0.148		0.82	0.55-1.22	0.192
Previous breast or axillary surgery	yes vs. no		0.76	0.50-1.16	0.093		0.85	0.57-1.27	0.308
Time to chemotherapy	Time-dependent variable		1.97	1.27-3.06	< 0.001		2.01	1.30-3.10	< 0.001
Time to early surgical complications	Time-dependent variable		1.84	1.23-2.73	< 0.001		1.73	1.18-2.55	< 0.001
Sexual functioning		162 (57)				168 (73)			
Type of surgery	ALD vs. LDI		1.36	0.83-2.24	0.112		1.17	0.75-1.82	0.364
Education	(> high school) vs. ≤ high school		1.43	0.88-2.32	0.058		1.16	0.75-1.79	0.387
Employed	yes vs. no		0.61	0.37-1.02	0.013		0.68	0.43-1.09	0.036
Time to chemotherapy	Time-dependent variable		1.94	1.16-3.25	0.001		1.94	1.20-3.14	< 0.001
Sexual enjoyment		98 (26)				120 (30)			
Type of surgery	ALD vs. LDI		0.90	0.45-1.82	0.702		0.96	0.49-1.85	0.862
Lymph node positivity	yes vs. no		1.93	0.96-3.88	0.016		1.98	1.02-3.87	0.009
Future perspective		166 (33)				170 (57)			
Type of surgery	ALD vs. LDI		0.67	0.36-1.28	0.112		0.66	0.40-1.07	0.026
Reason for surgery	symptomatic vs. screen-detected		3.01	1.45-6.22	< 0.001		1.96	1.19-3.23	0.001

LDI, implant-assisted latissimus dorsi breast reconstruction; ALD, autologous latissimus dorsi breast reconstruction; HR, hazard ratio; 99%CI, 99% confidence interval; Time to early surgical complications, up to 3 months after surgery; DCIS, Ductal carcinoma *in* situ; Education, ≤ high school includes the secondary school and before it, > high school includes the Advanced level and after it; AxLND, Axillary lymph nodes dissections.

Supplementary Table S4. Results of the multivariate Cox regression of the time to health-related quality of life score deterioration (TTD) for the symptomatic scales of the QLQ-C30 and QLQ-BR23 questionnaires.

		TTD as compared to the baseline score				TTD as compared to the best previous score			
		n (events)	**HR**	**99%CI**	**P**	**n (events)**	**HR**	**99%CI**	**P**
Fatigue		171 (102)				172 (116)			
Type of surgery	ALD vs. LDI		0.86	0.59-1.25	0.291		0.97	0.68-1.38	0.803
Anxiety					0.046				0.013
	Moderate vs. Normal		1.35	0.87-2.12	0.081		1.50	0.99-2.30	0.013
	Severe vs. Normal		1.47	0.96-2.25	0.020		1.48	0.98-2.24	0.014
Time to chemotherapy	Time-dependent variable		2.00	1.34-3.00	< 0.001		1.76	1.18-2.61	< 0.001
Time to early surgical complications	Time-dependent variable		1.36	0.95-1.94	0.028		1.38	0.98-1.95	0.015
Nausea and vomiting		171 (49)				172 (55)			
Type of surgery	ALD vs. LDI		0.66	0.39-1.11	0.038		0.73	0.44-1.19	0.096
Anxiety					0.003				0.006
	Moderate vs. Normal		0.79	0.36-1.76	0.452		1.10	0.55-2.19	0.728
	Severe vs. Normal		1.83	1.05-3.21	0.005		1.90	1.10-3.27	0.002
Time to chemotherapy	Time-dependent variable		3.21	1.89-5.43	< 0.001		2.79	1.68-4.63	< 0.001
Time to early surgical complications	Time-dependent variable		1.74	1.03-2.93	0.007		1.94	1.17-3.19	0.001
Pain		171 (89)				172 (96)			
Type of surgery	ALD vs. LDI		0.99	0.67-1.46	0.927		1.10	0.75-1.61	0.516
Previous breast or axillary surgery	yes vs. no		0.58	0.38-0.89	0.001		0.62	0.41-0.94	0.003
Anxiety					0.017				0.013
	Moderate vs. Normal		1.41	0.84-2.36	0.086		1.30	0.78-2.16	0.182
	Severe vs. Normal		1.61	1.04-2.49	0.006		1.63	1.06-2.48	0.003
Dyspnoea		171 (50)				172 (55)			
Type of surgery	ALD vs. LDI		1.34	0.77-2.32	0.173		1.20	0.71-2.01	0.372
Education	(> high school) vs. ≤ high school		0.55	0.32-0.96	0.006		0.57	0.34-0.96	0.005
Anxiety					< 0.001				< 0.001
	Moderate vs. Normal		2.53	1.24-5.18	0.001		2.19	1.14-4.24	0.002
	Severe vs. Normal		3.04	1.62-5.71	< 0.001		2.35	1.31-4.22	< 0.001
Time to chemotherapy	Time-dependent variable		2.11	1.18-3.79	0.001		2.13	1.23-3.71	< 0.001
Time to early surgical complications	Time-dependent variable		3.93	2.19-7.06	< 0.001		3.19	1.88-5.43	< 0.001
Insomnia		160 (64)				161 (74)			
Type of surgery	ALD vs. LDI		0.93	0.58-1.50	0.702		0.84	0.53-1.32	0.316
Education	(> high school) vs. ≤ high school		1.57	0.99-2.50	0.012		1.53	0.98-2.37	0.013
Reason for surgery	symptomatic vs. screen-detected		1.64	0.99-2.69	0.011		1.41	0.89-2.23	0.053
Type of axillary surgery	Sentinel node assisted sample vs. AxLND		0.45	0.24-0.87	0.002		0.58	0.33-1.03	0.014
Anxiety					0.002				0.021
	Moderate vs. Normal		1.09	0.64-1.85	0.685		1.29	0.78-2.13	0.191
	Severe vs. Normal		0.41	0.21-0.82	0.001		0.63	0.35-1.14	0.045
Time to early surgical complications	Time-dependent variable		1.02	0.63-1.65	0.935		1.10	0.69-1.75	0.619
Appetite loss		171 (39)				172 (45)			
Type of surgery	ALD vs. LDI		0.58	0.33-1.03	0.015		0.62	0.36-1.07	0.024
Type of tumour					< 0.001				0.020
	DCIS vs. Invasive		0.38	0.13-1.11	0.020		0.70	0.29-1.71	0.309
	Invasive+DCIS vs. Invasive		1.52	0.73-3.18	0.140		1.45	0.71-2.96	0.185
Time to early surgical complications	Time-dependent variable		2.31	1.27-4.20	< 0.001		2.36	1.34-4.15	< 0.001
Constipation		169 (51)				171 (56)			
Type of surgery	ALD vs. LDI		0.41	0.24-0.70	< 0.001		0.53	0.33-0.86	0.001
Previous breast or axillary surgery	yes vs. no		0.52	0.29-0.91	0.003		0.53	0.31-0.89	0.002
Type of tumour					< 0.001				0.004
	DCIS vs. Invasive		0.26	0.10-0.65	< 0.001		0.55	0.26-1.17	0.041
	Invasive+DCIS vs. Invasive		1.10	0.59-2.07	0.694		1.18	0.64-2.21	0.483
Anxiety					0.064				0.335
	Moderate vs. Normal		1.34	0.67-2.68	0.284		0.99	0.52-1.88	0.960
	Severe vs. Normal		1.67	0.95-2.95	0.019		1.32	0.78-2.24	0.170
Time to early surgical complications	Time-dependent variable		1.59	0.95-2.65	0.021		1.48	0.92-2.40	0.034
Diarrhoea		167 (31)				170 (39)			
Type of surgery	ALD vs. LDI		1.16	0.62-2.15	0.546		0.90	0.51-1.57	0.617
Reason for surgery	symptomatic vs. screen-detected		0.39	0.20-0.76	< 0.001		0.37	0.20-0.67	< 0.001
Time to early surgical complications	Time-dependent variable		3.38	1.73-6.61	< 0.001		2.71	1.51-4.87	< 0.001
Financial difficulties		166 (40)				170 (43)			
Type of surgery	ALD vs. LDI		0.69	0.38-1.24	0.100		0.82	0.47-1.43	0.362
Reason for surgery	symptomatic vs. screen-detected		2.31	1.20-4.45	0.001		2.37	1.27-4.43	< 0.001
Time to early surgical complications	Time-dependent variable		1.30	0.72-2.34	0.258		1.30	0.74-2.28	0.237
Systematic therapy side effects		169 (60)				172 (66)			
Type of surgery	ALD vs. LDI		0.70	0.44-1.13	0.056		0.68	0.43-1.08	0.032
Previous breast or axillary surgery	yes vs. no		0.72	0.43-1.21	0.103		0.83	0.52-1.34	0.316
Time to chemotherapy	Time-dependent variable		4.32	2.63-7.09	< 0.001		4.14	2.55-6.70	< 0.001
Time to early surgical complications	Time-dependent variable		1.70	1.06-2.71	0.004		1.53	0.98-2.39	0.015
Breast symptoms		165 (50)				171 (54)			
Type of surgery	ALD vs. LDI		0.63	0.38-1.05	0.020		0.68	0.42-1.10	0.039
Previous breast or axillary surgery	yes vs. no		0.69	0.40-1.19	0.077		0.73	0.44-1.24	0.128
Lymph node positivity	yes vs. no		1.76	1.06-2.90	0.004		1.69	1.04-2.75	0.006
Anxiety					0.006				0.008
	Moderate vs. Normal		0.53	0.25-1.12	0.029		0.58	0.29-1.18	0.050
	Severe vs. Normal		1.38	0.81-2.36	0.123		1.40	0.83-2.36	0.095
Time to early surgical complications	Time-dependent variable		1.29	0.79-2.12	0.180		1.14	0.71-1.84	0.490
Arm symptoms		165 (100)				171 (109)			
Type of surgery	ALD vs. LDI		0.79	0.54-1.16	0.116		0.79	0.54-1.14	0.096
Education	(> high school) vs. ≤ high school		1.38	0.96-2.01	0.023		1.34	0.93-1.91	0.038
Other medical illnesses	yes vs. no		0.66	0.36-1.22	0.081		0.59	0.32-1.10	0.029
Previous breast or axillary surgery	yes vs. no		0.36	0.23-0.55	< 0.001		0.41	0.28-0.62	< 0.001
Anxiety					0.004				0.005
	Moderate vs. Normal		1.27	0.79-2.03	0.199		1.38	0.88-2.15	0.064
	Severe vs. Normal		1.75	1.14-2.71	0.001		1.68	1.11-2.55	0.001
Time to chemotherapy	Time-dependent variable		1.35	0.89-2.06	0.063		1.37	0.90-2.07	0.051

LDI, implant-assisted latissimus dorsi breast reconstruction; ALD, autologous latissimus dorsi breast reconstruction; HR, hazard ratio; 99%CI, 99% confidence interval; Time to early surgical complications, up to 3 months after surgery; DCIS, Ductal carcinoma *in* situ; Education, ≤ high school includes the secondary school and before it, > high school includes the Advanced level and after it; AxLND, Axillary lymph nodes dissections; Other medical illnesses, other systematic conditions such as rheumatoid arthritis, clotting disorders, connective tissue diseases.

3.2. Essai de phase II randomisé chez des patients atteints d'un cancer du pancréas

<u>Article :</u> Impact on health-related quality of life deterioration-free survival of a first-line therapy combining nab-paclitaxel plus either gemcitabine or simplified leucovorin and fluorouracil for patients with metastatic pancreatic cancer: Results of the randomized phase II AFUGEM GERCOR clinical trial.

Article publié dans *Cancer Medicine*

Résumé

Objectif

L'essai de phase II AFUGEM GERCOR avait pour objectif d'évaluer l'efficacité d'une chimiothérapie de première ligne, associant du nab-paclitaxel et de la gemcitabine (groupe gemcitabine) ou de la leucovorine et du fluorouracile simplifiés (groupe LV5FU2 simplifié), chez des patients atteints d'un cancer du pancréas métastatique non-prétraité. Le critère de jugement principal de l'essai était la survie sans progression à 4 mois. Les résultats étaient en faveur du groupe LV5FU2 simplifié avec un taux de survie sans progression à 4 mois de 56% [IC90% 45-66] et de 54% [IC90% 40-68] pour le groupe gemcitabine. Cette étude portait sur l'analyse des données de QdV en tant qu'objectif secondaire, afin d'évaluer le bénéfice clinique pour le patient.

Méthodes

La QdV a été évaluée par le questionnaire EORTC QLQ-C30, à l'inclusion (avant la randomisation) et à chaque cycle de chimiothérapie jusqu'à la fin du traitement. Le profil des données de QdV manquantes à l'inclusion a été exploré. Les analyses ont été effectuées en comparant deux groupes de patients : les patients qui ont un score de QdV à l'inclusion disponible (population en intention de traiter modifiée) et ceux qui n'en ont pas. La survie sans détérioration de la QdV (QFS, « *health-related quality of life deterioration-free survival* ») a été utilisée comme modalité d'analyse longitudinale pour la population en intention de traiter modifiée. La QFS a été définie comme l'intervalle de temps entre l'inclusion du patient et l'apparition d'une première détérioration définitive du score de QdV

par rapport au score à l'inclusion, ou le décès. Des modèles de Cox univariés ont été réalisés pour estimer les HRs et les IC90% de l'effet du traitement. A titre exploratoire, des analyses multivariées de Cox ont également été réalisées afin d'étudier les facteurs influençant potentiellement la QFS. Plusieurs analyses de sensibilité ont été réalisées : tout d'abord, en excluant le décès comme événement; puis, en tenant compte de l'imputation des données manquantes à l'inclusion, afin de considérer tous les patients en intention de traiter. Les imputations multiples des items manquants à l'inclusion ont été effectuées selon la méthode de Monte-Carlo par chaînes de Markov, en tenant compte du profil des données manquantes. Des modèles de Cox multivariés ont ensuite été construits en introduisant les mêmes variables que pour l'analyse longitudinale avant le traitement des données manquantes.

Résultats

Entre 2013 et 2014, 114 patients ont été randomisés (ratio 1:2) : 39 patients dans le groupe gemcitabine et 75 patients dans le groupe LV5FU2 simplifié. Soixante-quatre patients (56,1%) ont complété le questionnaire QLQ-C30 à l'inclusion : 22 patients (56,4%) dans le groupe gemcitabine et 42 patients (56,0%) dans le groupe LV5FU2 simplifié. En ce qui concerne les caractéristiques à l'inclusion, les patients qui ont un score de QdV à l'inclusion disponible diffèrent des autres patients en termes d'aspartate aminotransférase (P = 0,037), de gamma-glutamyl transpeptidase (P = 0,086) et d'antigène carbohydrate 19-9 (P = 0,039). La médiane de survie globale était de 9,5 mois [IC90% 8,80-13,30] chez les patients sans score de QdV disponible à l'inclusion et de 11,2 mois [IC90% 8,57-15,80] chez les patients avec un score de QdV disponible à l'inclusion (HR = 1,08, [IC90% 0,76-1,53], P = 0,720). Ainsi, nous avons supposé que les données manquantes à l'inclusion dépendent des caractéristiques des patients à l'inclusion. Les patients du groupe LV5FU2 simplifié semblaient présenter une QFS plus longue que ceux du groupe gemcitabine pour 14 des 15 dimensions, avec des HRs < 1. Les résultats de l'analyse de sensibilité excluant le décès comme événement étaient significativement en faveur du groupe LV5FU2 simplifié pour les dimensions fonctionnement physique (HR = 0,51 [IC90% 0,27-0,97]) et douleur (HR = 0,26 [IC90% 0,09-0,74]). Après imputation des données manquantes à l'inclusion, les résultats des différentes analyses étaient similaires à ceux obtenus avant le traitement des données manquantes.

Conclusion

La combinaison des chimiothérapies nab-paclitaxel avec de la leucovorine et du fluorouracile simplifiés n'ont eu aucun impact négatif dans les analyses exploratoires de la QdV.

Received: 30 October 2018 | Revised: 10 May 2019 | Accepted: 15 May 2019

DOI: 10.1002/cam4.2311

ORIGINAL RESEARCH

Cancer Medicine Open Access WILEY

Impact on health-related quality of life deterioration-free survival of a first-line therapy combining nab-paclitaxel plus either gemcitabine or simplified leucovorin and fluorouracil for patients with metastatic pancreatic cancer: Results of the randomized phase II AFUGEM GERCOR clinical trial

Emilie Charton[1,2] | **Jean-Baptiste Bachet**[3] | **Pascal Hammel**[4] | **Jérôme Desramé**[5] | **Benoist Chibaudel**[6] | **Romain Cohen**[7] | **Philippe Debourdeau**[8] | **Jérome Dauba**[9] | **Thierry Lecomte**[10] | **Jean-François Seitz**[11] | **Christophe Tournigand**[12] | **Thomas Aparicio**[13] | **Véronique Guerin-Meyer**[14] | **Julien Taieb**[15] | **Julien Volet**[16] | **Christophe Louvet**[17] | **Amélie Anota**[1,2,18] | **Franck Bonnetain**[1,2,18]

[1]Methodology and Quality of Life Unit in Oncology, INSERM UMR 1098, University Hospital of Besançon, Besançon, France

[2]University Bourgogne Franche-Comté, INSERM, EFS BFC, UMR1098, Interactions Hôte-Greffon-Tumeur/Ingénierie Cellulaire et Génique, Besançon, France

[3]Department of Hepato-Gastroenterology, Groupe hospitalier Pitié Salpêtrière, Sorbonne University, UPMC University, Paris, France

[4]Department of Digestive Oncology, Hôpital Beaujon, Clichy, France

[5]Department of Hepato-Gastroenterology, Hôpital Privé Jean Mermoz, Lyon, France

[6]Department of Oncology, Institut Franco-Britannique, Levallois-Perret, France

[7]Department of Oncology, AP-HP, Hôpital Saint-Antoine, Sorbonne University, Paris, France

[8]Department of Oncology, Institut Saint Catherine, Avignon, France

[9]Department of Oncology, Hôpital Layne Mont de Marsan, Mont de Marsan, France

[10]Department of Hepato-Gastroenterology, Hôpital Trousseau, Tours, France

[11]CHU La Timone, Marseille, France

[12]Department of Oncology, CHU Henri Mondor, Créteil, France

[13]Department of Hepato-Gastroenterology, CHU Saint Louis, Paris, France

[14]Department of Oncology, Institut de cancérologie de L'Ouest Paul Papin, Angers, France

[15]Department of Gastroenterology and Digestive Oncology, Hôpital Européen Georges Pompidou, Paris, France

[16]Department of Hepato-Gastroenterology, CHU Robert Debré, Reims, France

[17]Department of Oncology, Institut Mutualiste Montsouris, Paris, France

[18]French National Platform of Quality of Life and Cancer, Besançon, France

Correspondence
Emilie Charton, Methodology and Quality of Life Unit in Oncology (INSERM UMR 1098), University Hospital of Besançon, 3 Boulevard Fleming, 25030 Besançon, France.
Email: echarton@chu-besancon.fr

Abstract

Background: The phase II AFUGEM GERCOR trial aimed to assess the efficacy of a first-line therapy combining nab-paclitaxel plus either gemcitabine (gemcitabine group) or simplified leucovorin and fluorouracil (sLV5FU2 group) in patients with

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2019 The Authors. *Cancer Medicine* published by John Wiley & Sons Ltd.

Funding information:
This work was supported by GERCOR (Groupe Coopérateur Multidisciplinaire en Oncologie), Celgene (through grants to GERCOR) and by a grant from the "Institut National du Cancer (INCA 11862)". The GERCOR collaborative group was involved in study design, data collection, data analysis, data interpretation, and writing of the report, and in the decision to submit for publication. Celgene provided funding to GERCOR, but had no role in any part of study design, data collection, data analysis, data interpretation, or writing of the report.

previously untreated metastatic pancreatic cancer. Results of progression-free survival at 4 months (primary endpoint) were in favor of the sLV5FU2 group. This paper presents health-related quality of life (HRQoL) data as a secondary endpoint.

Methods: HRQoL was assessed using the EORTC QLQ-C30 questionnaire at baseline and at each chemotherapy cycle until the end of treatment. The HRQoL deterioration-free survival (QFS) was used as a modality of longitudinal analysis. QFS was defined as the time between randomization and the first definitive HRQoL score deterioration as compared to the baseline score, or death. Sensitivity analysis was performed excluding death as an event. Univariate Cox models were used to estimate hazard ratios (HRs) and 90% confidence intervals (CIs) of the treatment effect.

Results: Between 2013 and 2014, 114 patients were randomized in a 1:2 ratio (39 in the gemcitabine group and 75 in the sLV5FU2 group). Patients in the sLV5FU2 group seemed to present longer QFS than those of the gemcitabine group for 14 out of 15 dimensions, with HRs < 1. Results of the sensitivity analysis excluding death as an event were significantly in favor of the sLV5FU2 group for physical functioning (HR = 0.51 [90% CI 0.27-0.97]) and pain (HR = 0.26 [90% CI 0.09-0.74]).

Conclusion: The nab-paclitaxel plus simplified leucovorin and fluorouracil combination had no negative impact in exploratory HRQoL analyses.

KEYWORDS
clinical trial, deterioration-free survival, metastatic, pancreatic cancer, quality of life

1 | INTRODUCTION

Pancreatic cancer is a devastating disease with an overall 5-year survival of less than 5%.[1,2] The mortality trend is increasing in both genders,[3,4] and pancreatic cancer is one of the most common causes of death from cancer.[5]

Over the past two decades, gemcitabine monotherapy has been a standard treatment for metastatic pancreatic cancer.[6] In 2011, the FOLFIRINOX regimen (fluorouracil, leucovorin, irinotecan, and oxaliplatin)[7,8] and the combination of gemcitabine with nab-paclitaxel[9,10] demonstrated an improvement in progression-free survival and overall survival compared with gemcitabine alone. These regimens are thus now considered as the standard first-line treatment options in patients with metastatic pancreatic cancer and good general status without marked cholestasis.

In the randomized, phase II, AFUGEM GERCOR (Groupe Coopérateur Multidisciplinaire en Oncologie) trial, nab-paclitaxel plus simplified leucovorin and fluorouracil treatment (sLV5FU2 group) improved the primary endpoint of progression-free survival at 4 months in the first 72 assessable patients in the sLV5FU2 group, and the secondary endpoint of overall survival compared to nab-paclitaxel plus gemcitabine treatment (gemcitabine group).[11] At 4 months, 40 (56% [90% confidence interval (CI) 45-66]) out of 72 patients in the sLV5FU2 group were alive and free of disease progression vs 21 (54% [90% CI 40-68]) out of 39 patients in the gemcitabine group. In exploratory analyses, the median progression-free survival was 5.9 months [95% CI 3.6-7.4] in the sLV5FU2 group vs 4.9 months [95% CI 2.1-7.7] in the gemcitabine group. Similarly, the median overall survival was 11.4 months [95% CI 8.8-16.5] in the sLV5FU2 group vs 9.2 months [95% CI 6.0-13.6] in the gemcitabine group (exploratory hazard ratio (HR) of 0.61 [95% CI 0.40-0.95]). Although these results appear to be promising for the nab-paclitaxel plus sLV5FU2 combination, it is crucial to study the impact of the treatment on patients' health-related quality of life (HRQoL) over time. In fact, new combinations of drugs can cause adverse events that may deteriorate the patients' perception of their health. Thus, it is particularly important to take the patient's HRQoL level into account in disease management, in order to ensure that the new treatment does not yield a clinical benefit at the cost of reduced quality of life.

In this context, based on the phase II AFUGEM clinical trial, we report the impact on HRQoL of a first-line therapy combining nab-paclitaxel plus either gemcitabine or sLV5FU2 in patients with previously untreated metastatic pancreatic cancer.

2 | METHODS

2.1 | Patients and eligibility criteria

The AFUGEM study was an open-label, noncomparative, randomized, multicentre, phase II clinical trial,

conducted in 15 centers in France (ClinicalTrials.gov number NCT01964534).

Eligible patients were required to be aged at least 18 years, with histologically or cytologically proven adenocarcinoma of the pancreas, stage IV disease, no prior therapy for metastatic disease, an Eastern Cooperative Oncology Group (ECOG) performance status 0-2, and presenting adequate hematologic, renal, and liver function. The detailed eligibility criteria have previously been reported.[12] The protocol was approved by the French Ethics Committee and written informed consent was obtained from all patients before randomization.

Using a minimization technique stratified by center and ECOG performance status, patients were randomly assigned (1:2 ratio) to receive nab-paclitaxel plus gemcitabine (control arm) or nab-paclitaxel plus sLV5FU2 (experimental arm). Both regimens were administered every 28 days and details of the regimens have previously been published.[12]

The primary endpoint was progression-free survival at 4 months in the first 72 patients in the sLV5FU2 group. Secondary endpoints were objective response, progression-free survival, overall survival, tolerance and HRQoL.

2.2 | Health-related quality of life assessment

HRQoL was assessed in each treatment arm using the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 (EORTC QLQ-C30) cancer-specific questionnaire,[13] at baseline and at each chemotherapy cycle until the end of treatment. The QLQ-C30 includes 30 items and assesses global health status, 5 functional scales (physical, role, emotional, cognitive and social functioning) and 9 symptom scales (fatigue, nausea and vomiting, pain, dyspnoea, insomnia, appetite loss, constipation, diarrhea, and financial difficulties). Scores vary from 0 (worst) to 100 (best) for global health status and functional scales, and from 0 (best) to 100 (worst) for the symptom scales.

2.3 | Statistical analysis

2.3.1 | Population and statistical considerations

The intention-to-treat (ITT) population was considered in the HRQoL analysis, that is, all randomized patients regardless of their eligibility criteria and treatment received. Due to the occurrence of missing data, a modified ITT (mITT) population was also defined as all ITT patients with at least one HRQoL score available at baseline.[14]

Five targeted dimensions were defined a priori in the protocol: physical functioning, emotional functioning, fatigue, pain, and appetite loss. Other dimensions were regarded as being exploratory.[12]

Although HRQoL was a secondary endpoint in this study, a decision-rule was integrated into the protocol to facilitate interpretation of the results, as follows: HRQoL would be considered as being improved in one arm if at least one time to HRQoL score deterioration among the 5 targeted dimensions was significantly longer without a significantly shorter time to HRQoL score deterioration for the other 4 targeted dimensions.

Since AFUGEM study is a noncomparative study, *P*-values of the treatment effect are not reported, while effect sizes are presented for exploratory purposes using HRs and 90% CIs. A 5 point difference in HRQoL scores was considered as the minimal clinically important difference.[15]

2.3.2 | Descriptive analysis at baseline

Quantitative variables are described using median and range. Qualitative variables are summarized using number and percentage.

The profile of missing HRQoL data at baseline was explored.[16] Analyses were carried out by comparing 2 groups of patients: patients who completed the baseline HRQoL questionnaire (mITT population) versus those who did not. In order to determine whether baseline missing data were dependent on the patients' characteristics, the comparison was performed according to baseline clinical and socio-demographic variables, using the *t* test or Mann-Whitney nonparametric test for continuous variables, and χ^2 or Fisher's exact test for qualitative variables. *P*-values < 0.1 were considered as significant. To determine whether baseline missing data depended on patients' health status, the comparison was performed according to overall survival. Overall survival curves were estimated using the Kaplan-Meier estimation method, described using median and 90% CI and compared using the log-rank test. Univariate Cox analysis was used to estimate the HR and 90% CI.

2.3.3 | Longitudinal analysis

HRQoL deterioration-free survival (QFS) was used as a modality of longitudinal analysis. QFS was defined as the time between randomization and the first HRQoL score deterioration of at least 5 points, as compared to the baseline score, with no further improvement of at least 5 points as compared to the baseline score, or death, whichever occurred first.[17,18] QFS curves were estimated using the Kaplan-Meier estimation method. Univariate Cox models were used to calculate HRs and 90% CIs of the treatment effect. All variables collected at baseline were tested by univariate Cox analysis. The impact of time to toxicity grade 3-4 was also tested by univariate analysis as a time-dependent variable.

Variables significant at a threshold of 10% by univariate analyses were eligible for the multivariate model. The treatment arm was forced in the model. Restricted mean method was used as a supplement to the HR in case of nonrespect of the proportional hazards assumption in the Cox model, and also to ensure the robustness of the model. The difference of restricted mean survival time was estimated with 90% CI. A difference of restricted mean survival time greater than zero favored the sLV5FU2 group.[19]

2.3.4 | Sensitivity analyses

Several sensitivity analyses were performed.

First, analyses were repeated excluding death as an event in the QFS definition, which then becomes simply the time until definitive HRQoL deterioration (TUDD).[17,18]

Then, QFS analyses were repeated after imputation of baseline missing data in order to consider all ITT patients. Multiple imputations of baseline missing items were performed using the Markov Chain Monte Carlo method taking into account the profile of missing data. Multivariate models were then constructed introducing the same variables as for the longitudinal analysis before treatment of missing data.

Analyses were performed using SAS (version 9.3) (SAS Institute Inc, Cary, NC) and R (version 3.3.1) software.

3 | RESULTS

3.1 | Study population

Between December 2013 and October 2014, 114 patients were randomized: 39 in the gemcitabine group and 75 in the sLV5FU2 group. Sixty-four patients (56.1%) completed the QLQ-C30 questionnaire at baseline, 22 patients (56.4%) in the gemcitabine group, and 42 (56.0%) in the sLV5FU2 group (Figure 1). Patients in the sLV5FU2 group completed the questionnaire longer after compared to the gemcitabine group. Indeed, there were more patients included in this treatment arm due to the ratio 1:2 and the median overall survival was significantly longer in this treatment arm. The median age was 66 years (range 45-86) and 70 patients (61.4%) were men. The baseline characteristics of the patients are described according to baseline HRQoL availability in Table 1. Baseline HRQoL level was similar between treatment arms (Table S1).

3.2 | Missing data analysis

Regarding the baseline characteristics, patients with available baseline HRQoL differed from other patients in terms of aspartate aminotransferase ($P = 0.037$), gamma-glutamyl transpeptidase ($P = 0.086$) and cancer antigen 19-9

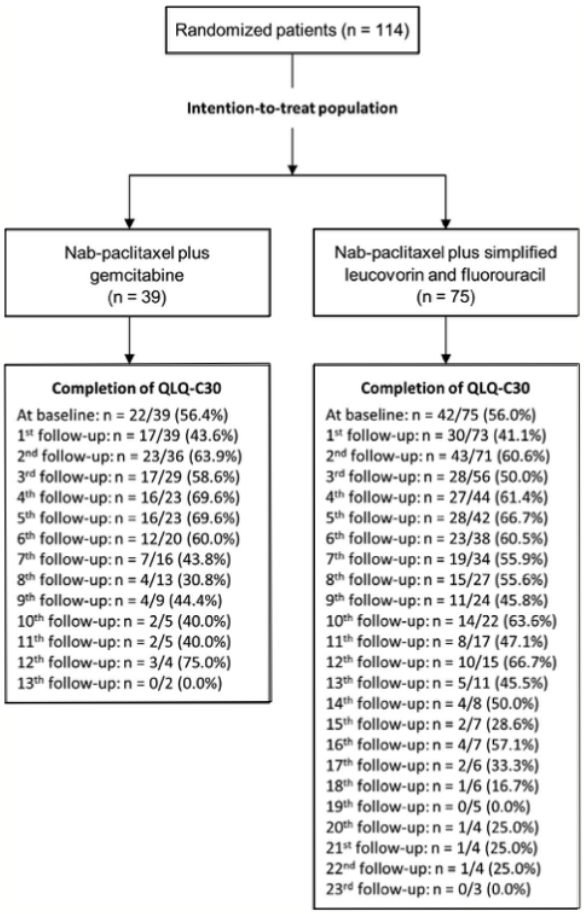


FIGURE 1 CONSORT diagram for EORTC QLQ-C30 questionnaire

($P = 0.039$) (Table 1). The median overall survival was 9.5 months [90% CI 8.80-13.30] for patients without available baseline HRQoL vs 11.2 months [90% CI 8.57-15.80] for mITT patients (HR = 1.08, [90% CI 0.76-1.53], $P = 0.720$) (Figure 2). Thus, we could suppose that missing data at baseline depend only on patients' baseline characteristics.

3.3 | Longitudinal analysis

Analyses of the mITT population showed a trend towards longer QFS in favor of the sLV5FU2 group among the five targeted dimensions, namely physical functioning (HR = 0.64 [90% CI 0.40-1.03]), emotional functioning (HR = 0.71 [90% CI 0.44-1.16]), fatigue (HR = 0.79 [90% CI 0.50-1.26]), pain (HR = 0.62 [90% CI 0.38-1.01]) and appetite loss (HR = 0.70 [90% CI 0.44-1.13]), with HRs < 1 (Figure 3). Similar trends were observed for all the other dimensions, except for constipation, which had a HR of 1.04 [90% CI 0.64-1.69] (Table S2).

TABLE 1 Baseline characteristics of patients according to the availability of their baseline health-related quality of life (HRQoL) questionnaire

	Baseline HRQoL available (n = 64)		HRQoL not available (n = 50)		
	n	%	n	%	*P*
Gender					
Male	38	59.4	32	64.0	0.615
Women	26	40.6	18	36.0	
Arm					
Gemcitabine	22	34.4	17	34.0	0.967
sLV5FU2	42	65.6	33	66.0	
ECOG Perormance Status					
0	24	37.5	13	26.0	0.193
1, 2	40	62.5	37	74.0	
Pain					
Yes	36	56.3	21	42.0	0.131
No	28	43.7	29	58.0	
Jaundice					
Yes	2	3.1	4	8.0	0.402[c]
No	62	96.9	46	92.0	
Ascites					
Yes	4	6.3	2	4.0	0.694[c]
No	60	93.7	48	96.0	
Hemoglobin					
Normal	39	60.9	30	60.0	0.919
Abnormal[a]	25	39.1	20	40.0	
Platelets					
Normal	52	81.3	44	88.0	0.327
Abnormal[a]	12	18.7	6	12.0	
Total bilirubin					
Normal	52	81.3	41	82.0	0.918
Abnormal[b]	12	18.7	9	18.0	
Alkaline phosphatase					
Normal	31	49.2	20	40.8	0.376
Abnormal[b]	32	50.8	29	59.2	
Aspartate aminotransferase					
Normal	51	79.7	31	62.0	0.037
Abnormal[b]	13	20.3	19	38.0	
Alanine aminotransferase					
Normal	45	70.3	33	66.0	0.623
Abnormal[b]	19	29.7	17	34.0	
Albumin					
Normal	39	60.9	36	73.5	0.162
Abnormal[a]	25	39.1	13	26.5	
Carcinoembryonic antigen					

(Continues)

TABLE 1 (Continued)

	Baseline HRQoL available (n = 64)		HRQoL not available (n = 50)		
	n	%	n	%	*P*
Normal	23	38.3	20	46.5	0.407
Abnormal[b]	37	61.7	23	53.5	
Age (years)[f]	64	65.6 (47-85)	50	66.4 (45-86)	0.702[d]
Body mass index (kg/m^2)[f]	64	23.0 (16-33)	50	23.5 (16-32)	0.610[d]
Neutrophils (/mm^3)[f]	64	5425.0 (2592-19704)	50	5707.0 (1624-12168)	0.690[e]
Creatinine (µmol/L)[f]	64	66.5 (42-135)	50	66.2 (29-140)	0.671[e]
Gamma-glutamyl transpeptidase (U/L)[f]	61	134.0 (14-1564)	44	204.5 (15-920)	0.086[e]
Cancer antigen 19-9 (UI/L)[f]	61	475.2 (2-481206)	44	1391.0 (5-214000)	0.039[e]

A χ^2 test is used unless indicated otherwise.
[a]abnormally low or high levels
[b]abnormally high level
[c]Fisher's exact test
[d]*t* test
[e]Mann-Whitney nonparametric test; sLV5FU2, simplified leucovorin and fluorouracil
[f]Median (range) for continuous variables.

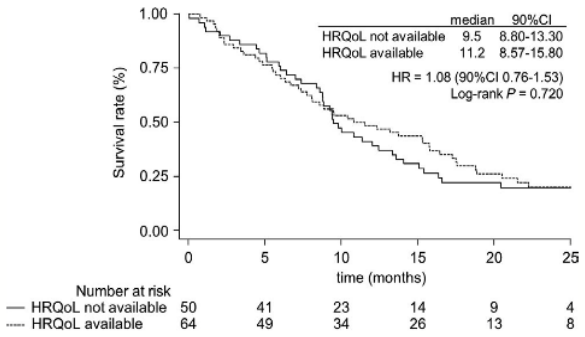


FIGURE 2 Overall survival curves according to baseline health-related quality of life (HRQoL) availability. HR, hazard ratio; 90% CI, 90% confidence interval

Multivariate analyses showed significantly shorter QFS among patients with the following (Table S3):

- An abnormally low level of albumin for emotional functioning (HR = 2.50 [90% CI 1.49-4.18]) and appetite loss (HR = 2.59 [90% CI 1.50-4.46]),
- other symptoms for fatigue (HR = 1.75 [90% CI 1.01-3.06]),
- abnormally low level of hemoglobin for pain (HR = 3.36 [90% CI 1.98-5.72]),
- abnormally high level of aspartate aminotransferase for appetite loss (HR = 1.86 [90% CI 1.01-3.45]).

Results also showed significantly longer QFS among patients with the following:

- An abnormally high level of carcinoembryonic antigen for physical functioning (HR = 0.50 [90% CI 0.28-0.88]),
- creatinine level greater than 66.2 µmol/L for fatigue (HR = 0.61 [90% CI 0.38-0.98]),
- included in the sLV5FU2 group for pain (HR = 0.54 [90% CI 0.32-0.92]) and appetite loss (HR = 0.55 [90% CI 0.33-0.91]).

3.4 | Sensitivity analysis

In the sensitivity analysis excluding death as an event, TUDD results were significantly in favor of the sLV5FU2 group for physical functioning (HR = 0.51 [90% CI 0.27-0.97]) and pain (HR = 0.26 [90% CI 0.09-0.74]) (Table S4). A trend toward longer TUDD in favor of the sLV5FU2 group among all other dimensions was observed, except for constipation.

After imputation of baseline missing data, results of the QFS analysis by treatment arm were similar to those obtained in the mITT population (Table S2).

Multivariate analyses after multiple imputations were in accordance with those obtained in the mITT population (Table S3). Results showed significantly shorter QFS among patients with the following:

- An abnormally high level of aspartate aminotransferase for emotional functioning (HR = 1.50 [90% CI 1.01-2.23]),

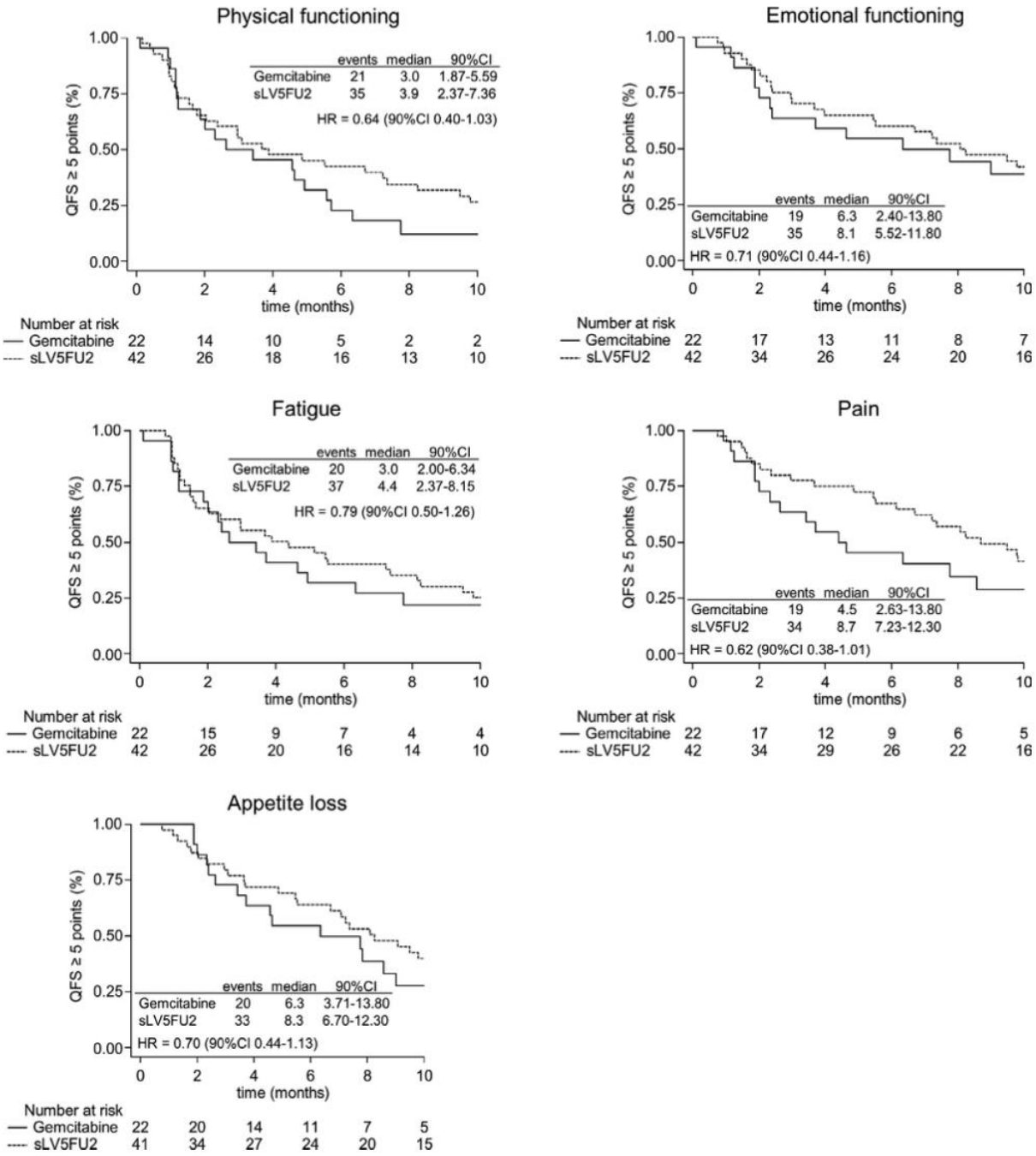


FIGURE 3 Kaplan-Meier curves of health-related quality of life deterioration-free survival (QFS). HR, hazard ratio; 90% CI, 90% confidence interval

- abnormally low level of albumin for emotional functioning (HR = 2.38 [90% CI 1.60-3.54]) and appetite loss (HR = 2.04 [90% CI 1.38-3.02]),
- abnormally low level of hemoglobin for fatigue (HR = 1.47 [90% CI 1.02-2.11]) and pain (HR = 2.19 [90% CI 1.48-3.23]),
- neutrophils level greater than 5590.0/mm^3 for pain (HR = 1.63 [90% CI 1.14-2.34]).

Results also showed significantly longer QFS among patients with a creatinine level greater than 66.2 µmol/L for fatigue (HR = 0.60 [90% CI 0.41-0.89]).

4 | DISCUSSION

In this QFS analysis performed on the population of the AFUGEM study, there was no negative impact on HRQoL in patients who received the combination of nab-paclitaxel plus sLV5FU2. When excluding death as an event, we found that this combination improved the patients' HRQoL level compared to gemcitabine group in terms of the decision-rule. In fact, among the 5 targeted dimensions, TUDD was significantly longer for 2 dimensions, namely physical functioning (HR = 0.51 [90% CI 0.27-0.97]) and pain (HR = 0.26 [90% CI 0.09-0.74]), without a significantly shorter TUDD for the other three targeted dimensions.

In the NAPOLI-1 phase III clinical trial in metastatic pancreatic cancer also comparing different combinations of chemotherapy, reported that liposomal irinotecan plus 5-fluorouracil and leucovorin maintained HRQoL as assessed with the QLQ-C30 versus 5-fluorouracil and leucovorin while improved overall survival.[20] However, this trial was performed in second-line treatment and thus is not completely similar to the AFUGEM study. In another phase III trial comparing gemcitabine to the PEFG (cisplatin, epirubicin, 5-fluorouracil, gemcitabine) regimen suggested that the combination improved overall survival while providing more grade 3-4 neutropenia and thrombocytopenia.[21] The impact of the treatment on HRQoL was thus important to study and the authors reported a maintained HRQoL using also the QLQ-C30 among other questionnaires.[22] Thus, our results are consistent with other studies using similar chemotherapy combinations.

Moreover, these results are consistent with the tolerable toxicity profile of the combination nab-paclitaxel plus sLV5FU2 highlighted previously.[11] These physician-reported toxicities are thus in accordance with the patients' perception of their own HRQoL. However, only general symptoms related to cancer were assessed using the QLQ-C30 questionnaire and specific pancreatic cancer symptoms, such as altered bowel habits or indigestion symptoms, could not be captured. Although the EORTC pancreatic cancer module was available at the time of the study,[23] it was not administered in this phase II trial to limit the patient burden and thus, occurrence of missing data.

The main limitation of this study was the relatively low proportion (56.1%) of questionnaires completed at baseline. However, the completion rate was similar in both treatment arms. A likely explanation could be that the first two HRQoL assessment times were at very short interval (ie, at randomization and day one of the first chemotherapy cycle) and this may have caused some confusion between these two time points at the time of data collection. The completion rate over time was close to those observed in other clinical trials in pancreatic cancer, such as in a phase II clinical trial in resectable or borderline resectable pancreatic cancer patients using also the QLQ-C30 questionnaire to assess HRQoL among other questionnaires.[24] These missing data also had a direct impact on the population for analysis. It was required to define a modified ITT population including all ITT patients with a baseline HRQoL score available. However, it is recommended that the ITT population be studied in order to best reflect the treatment effect without inducing bias.[14] It is therefore essential to verify that the modified ITT is representative of the ITT population.

We addressed this issue in the sensitivity analysis, using multiple imputations. The profile of missing data was explored and was dealt only at baseline, and we hypothesized that missing data depended on the baseline characteristics of the patients themselves. Our results showed a trend toward longer QFS in favor of the sLV5FU2 group for 14 out of 15 dimensions, in analyses before and after treatment of missing data. Thus, the occurrence of missing data does not appear to have biased the results. In fact, determining the mechanism of missing data is of fundamental importance to identify the appropriate strategy for analysis of these missing data,[25,26] but this remains rare in the analysis of HRQoL in oncology clinical trials.[27] Inadequate consideration and handling of missing data in the analysis can bias the results.[25,28] In the future, although it is a wellknown fact, it remains important to find the necessary means in the HRQoL data collection during the study to avoid missing data. HRQoL is now recognized as a key endpoint and decision criterion, and should be assessed as rigorously as other, classical biological and clinical endpoints that are required in clinical trials.

Another important point was the consideration of death in the time to HRQoL deterioration method. In this study, two approaches were adopted: first, excluding death, and second, considering death as an event within definition of time to HRQoL deterioration. Since a large number of deaths were observed, considering death as an event seemed to be the most appropriate solution and thus was retained as our main analysis. However, future research must pay greater attention to the consideration of death, for example by exploring the competing risk between death and deterioration of HRQoL.

In conclusion, this study suggests that nab-paclitaxel plus simplified leucovorin and fluorouracil does not have a negative impact on HRQoL compared to nab-paclitaxel plus gemcitabine. Thus, this combination of chemotherapy does not yield a clinical benefit at the cost of reduced quality of life.

ACKNOWLEDGMENTS

The authors thank Fiona Ecarnot (EA3920, University Hospital Besancon, University of Franche-Comté, Besançon, France) for editorial assistance.

CONFLICT OF INTEREST

J-BB has received personal fees from Amgen, Bayer, Celgene, Merck Serono, Roche, Sanofi, Servier, and nonfinancial support from Amgen, Merck Serono, and Roche. PH has received grants from Celgene and Roche; personal fees from Baxalta, Celgene, Ipsen, Lilly, Merck Serono, Novartis, and Pfizer; and nonfinancial support from Celgene, Ipsen, Merck Serono, Novartis, and Pfizer. BC has received personal fees from Bayer, Kantar Health, Kephren, Lilly, Sanofi, and nonfinancial support from Amgen, Merck Serono, and Roche. PD has received grants from DRCD Paris and personal fees from ITAC CME. J-FS has received personal fees from Bayer, Celgene, Lilly, Novartis, Pfizer, and Servier, and nonfinancial support from Roche. JT has received personal fees from Amgen, Celgene, Baxalta, Merck Serono, Sanofi, Sirtex, and Roche. CL received personal fees from Celgene, Roche, and Sanofi. FB received grants from Novartis and Roche; personal fees from BMS, Celgene, Integragen, Ipsen, Janssen, Merck Serono, Nestle, Novartis, and Roche; and nonfinancial support from Celgene, Ipsen, Merck Serono, Novartis, and Roche. All other authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Emilie Charton: Formal analysis, writing – original draft, and writing – review and editing. Jean-Baptiste Bachet: Conceptualization, data curation, funding acquisition, resources, supervision, validation, writing – original draft, and writing – review and editing. Pascal Hammel: Conceptualization, resources, and writing – review and editing. Jérôme Desramé: Resources, and writing – review and editing. Benoist Chibaudel: Resources, and writing – review and editing. Romain Cohen: Writing – review and editing. Philippe Debourdeau: Resources, and writing – review and editing. Jérome Dauba: Resources, and writing – review and editing. Thierry Lecomte: Resources, and writing – review and editing. Jean-François Seitz: Resources, and writing – review and editing. Christophe Tournigand: Resources, and writing – review and editing. Thomas Aparicio: Resources, and writing – review and editing. Véronique Guerin-Meyer: Resources, and writing – review and editing. Julien Taieb: Resources, and writing – review and editing. Julien Volet: Resources, and writing – review and editing. Christophe Louvet: Conceptualization, resources, and writing – review and editing. Amélie Anota: Formal analysis, supervision, validation, writing – original draft, and writing – review and editing. Franck Bonnetain: Conceptualization, methodology, resources, supervision, validation, writing – original draft, and writing – review and editing.

DATA AVAILABILITY STATEMENT

Data are unsuitable for public deposition due to ethical and legal restrictions and are therefore available upon request with the signature of a data privacy form. To request the data, the readers may contact Prof. Jean-Baptiste Bachet (jean-baptiste.bachet@aphp.fr).

ORCID

Emilie Charton https://orcid.org/0000-0003-0259-8811
Amélie Anota https://orcid.org/0000-0001-6896-1367

REFERENCES

1. Cooperman AM. Cancer of the pancreas: a dilemma in treatment. *Surg Clin North Am*. 1981;61(1):107-115.
2. Carpelan-Holmström M, Nordling S, Pukkala E, et al. Does anyone survive pancreatic ductal adenocarcinoma? A nationwide study re-evaluating the data of the Finnish Cancer Registry. *Gut*. 2005;54(3):385-387.
3. Malvezzi M, Bertuccio P, Levi F, La Vecchia C, Negri E. European cancer mortality predictions for the year 2014. *Ann Oncol Off J Eur Soc Med Oncol*. 2014;25(8):1650-1656.
4. Malvezzi M, Bertuccio P, Rosso T, et al. European cancer mortality predictions for the year 2015: does lung cancer have the highest death rate in EU women? *Ann Oncol Off J Eur Soc Med Oncol*. 2015;26(4):779-786.
5. Ferlay J, Colombet M, Soerjomataram I, et al. Cancer incidence and mortality patterns in Europe: Estimates for 40 countries and 25 major cancers in 2018. *Eur J Cancer Oxf Engl* 2018;103:356–387.
6. Burris HA, Moore MJ, Andersen J, et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. *J Clin Oncol Off J Am Soc Clin Oncol*. 1997;15(6):2403–2413.
7. Conroy T, Desseigne F, Ychou M, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N Engl J Med*. 2011;364(19):1817–1825.
8. Gourgou-Bourgade S, Bascoul-Mollevi C, Desseigne F, et al. Impact of FOLFIRINOX compared with gemcitabine on quality of life in patients with metastatic pancreatic cancer: results from the PRODIGE 4/ACCORD 11 randomized trial. *J Clin Oncol Off J Am Soc Clin Oncol*. 2013;31(1):23–29.
9. Von Hoff DD, Ervin T, Arena FP, et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. *N Engl J Med*. 2013;369(18):1691–1703.
10. Goldstein D, El-Maraghi RH, Hammel P, et al. nab-Paclitaxel plus gemcitabine for metastatic pancreatic cancer: long-term survival from a phase III trial. *J Natl Cancer Inst*. 2015;107(2):dju413–dju413.
11. Bachet J-B, Hammel P, Desramé J, et al. Nab-paclitaxel plus either gemcitabine or simplified leucovorin and fluorouracil as first-line therapy for metastatic pancreatic adenocarcinoma (AFUGEM GERCOR): a non-comparative, multicentre, open-label, randomised phase 2 trial. *Lancet Gastroenterol Hepatol*. 2017;2(5):337–346.

12. Bachet J-B, Chibaudel B, Bonnetain F, et al. A randomized phase II study of weekly nab-paclitaxel plus gemcitabine or simplified LV5FU2 as first-line therapy in patients with metastatic pancreatic cancer: the AFUGEM GERCOR trial. *BMC Cancer*. 2015;15:653.
13. 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(5):365–376.
14. Bonnetain F, Fiteni F, Efficace F, Anota A. Statistical Challenges in the Analysis of Health-Related Quality of Life in Cancer Clinical Trials. *J Clin Oncol*. 2016;34(16):1953–1956.
15. Osoba D, Rodrigues G, Myles J, Zee B, Pater J. Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol Off J Am Soc Clin Oncol*. 1998;16(1):139–144.
16. Little RJ, Rubin DB. *Statistical analysis with missing data*. John Wiley & Sons; 2002.
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*. 2010;46(15):2753–2762.
18. 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*. 2015;24(1):5–18.
19. Trinquart L, Jacot J, Conner SC, Porcher R. Comparison of Treatment Effects Measured by the Hazard Ratio and by the Ratio of Restricted Mean Survival Times in Oncology Randomized Controlled Trials. *J Clin Oncol Off J Am Soc Clin Oncol*. 2016;34(15):1813–1819.
20. Hubner RA, Cubillo A, Blanc J-F, et al. Quality of life in metastatic pancreatic cancer patients receiving liposomal irinotecan plus 5-fluorouracil and leucovorin. *Eur J Cancer Oxf Engl*. 1990;2019(106):24–33.
21. Reni M, Cordio S, Milandri C, et al. Gemcitabine versus cisplatin, epirubicin, fluorouracil, and gemcitabine in advanced pancreatic cancer: a randomised controlled multicentre phase III trial. *Lancet Oncol*. 2005;6(6):369–376.
22. Reni M, Bonetto E, Cordio S, et al. Quality of life assessment in advanced pancreatic adenocarcinoma: results from a phase III randomized trial. *Pancreatology*. 2006;6(5):454–463.
23. Fitzsimmons D, Johnson CD, George S, et al. Development of a disease specific quality of life (QoL) questionnaire module to supplement the EORTC core cancer QoL questionnaire, the QLQ-C30 in patients with pancreatic cancer. EORTC Study Group on Quality of Life. *Eur J Cancer Oxf Engl*. 1999;35(6):939–941.
24. Serrano PE, Herman JM, Griffith KA, et al. Quality of life in a prospective, multicenter phase 2 trial of neoadjuvant full-dose gemcitabine, oxaliplatin, and radiation in patients with resectable or borderline resectable pancreatic adenocarcinoma. *Int J Radiat Oncol Biol Phys*. 2014;90(2):270–277.
25. Fielding S, Fayers PM, Ramsay CR. Investigating the missing data mechanism in quality of life outcomes: a comparison of approaches. *Health Qual Life Outcomes*. 2009;7:57.
26. Fielding S, Ogbuagu A, Sivasubramaniam S, MacLennan G, Ramsay CR. Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. 2016;25(12):2977–2983.
27. Fiteni F, Anota A, Westeel V, Bonnetain F. Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a critical review. *BMC Cancer*. 2016;16:122.
28. Fielding S, Maclennan G, Cook JA, Ramsay CR. A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. *Trials*. 2008;9:51.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Charton E, Bachet J-B, Hammel P, et al. Impact on health-related quality of life deterioration-free survival of a first-line therapy combining nab-paclitaxel plus either gemcitabine or simplified leucovorin and fluorouracil for patients with metastatic pancreatic cancer: Results of the randomized phase II AFUGEM GERCOR clinical trial. *Cancer Med*. 2019;00:1–10. https://doi.org/10.1002/cam4.2311

Discussion et perspectives

En cancérologie, la QdV est un critère de jugement majeur pour évaluer le bénéfice clinique de nouvelles stratégies de traitement pour les patients (Osoba, 2011). L'analyse longitudinale des données de QdV reste complexe et non standardisée, ce qui compromet la comparaison des résultats entre les essais cliniques (Bonnetain et al., 2016; Bottomley et al., 2016). Diverses méthodes ont été proposées pour l'analyse longitudinale des données de QdV. Le temps jusqu'à détérioration est de plus en plus utilisé en tant que méthode d'analyse longitudinale dans les essais cliniques de phase III en cancérologie. La flexibilité de cette méthode permet l'adaptation à différents contextes, notamment à la situation thérapeutique. Malgré les premières recommandations concernant les définitions du temps jusqu'à détérioration à utiliser, l'absence de consensus reste une limite majeure à son utilisation. Cette approche mérite donc une attention particulière pour améliorer les recommandations actuelles.

Au cours de ce doctorat, mes différents travaux ont été centrés sur la méthode du temps jusqu'à détérioration afin d'apporter des éléments de discussion autour de sa standardisation.

Un premier objectif de ce travail a été de réaliser une revue systématique de la littérature sur la méthode du temps jusqu'à détérioration. Cette revue avait pour objectif d'évaluer la manière dont cette approche a été définie et rapportée dans les essais cliniques randomisés de phase III depuis 2014 pour analyser les données de PROs. En particulier, le niveau de clarté de la définition utilisée a été examiné dans les études par le biais de cinq critères clés. La concordance a également été étudiée entre les résultats des données de PROs analysés selon l'approche du temps jusqu'à détérioration et les résultats du (des) critère(s) de jugement principal (principaux) de l'étude.

Les résultats ont montré que le niveau de clarté de la définition utilisée était globalement bon. Ce niveau de clarté était plus élevé lorsque les PROs étaient reportés dans une publication dédiée. Ce résultat n'est pas surprenant puisque davantage d'informations autour de la méthodologie utilisée peuvent être reportées. Inversement, lorsque la QdV est un critère secondaire ou exploratoire de l'étude, l'espace dédié pour les résultats de QdV dans une publication principale reste très limité. Ainsi, il semble pertinent d'encourager l'écriture d'articles dédiés aux données de PROs, dans le but d'améliorer la qualité des résultats rapportés. Malgré tout, les chercheurs doivent faire un effort pour ne pas publier ces résultats

trop tardivement par rapport à la publication principale.

Les cinq critères clés permettant de qualifier le temps jusqu'à détérioration devraient être clairement définis dans la méthodologie et de façon systématique. Parmi ces critères, la façon de traiter les données manquantes à l'inclusion était le paramètre le moins bien défini quel que soit le type de publication. Il est primordial de savoir comment les données manquantes à l'inclusion sont prises en compte afin de pouvoir appliquer la méthode du temps jusqu'à détérioration.

Une concordance entre les résultats de l'approche du temps jusqu'à détérioration et le(s) critère(s) de jugement principal (principaux), a été mise en évidence pour la majorité des études incluses (n = 27, 69,2%). Cependant, il est important de préciser qu'une cohérence parfaite n'était pas attendue entre ces critères. En effet, un traitement expérimental efficace peut entraîner des toxicités qui nuiront à la QdV des patients. Il aurait donc été intéressant, dans cette revue, d'étudier les résultats de toxicités reportés par les cliniciens afin de pouvoir entièrement apprécier la cohérence des données de QdV au regard des autres résultats de l'étude.

Cette revue systématique de la littérature a été limitée à 4 années, de 2014 à 2018, en raison de la publication en 2013 des recommandations initiales sur l'approche du temps jusqu'à détérioration (Anota et al., 2013). De façon générale, cette revue a montré que peu d'études tenaient compte de ces recommandations. Diverses raisons peuvent expliquer ce manque de considération. Tout d'abord, la communication autour de cet article a pu être insuffisante. Les chercheurs peuvent considérer que ces recommandations ne sont pas adaptées à toute localisation cancéreuse ou situation thérapeutique. De plus, les plans d'analyses statistiques des essais ont possiblement été rédigés avant la publication des recommandations en 2013. La difficulté de réalisation des analyses peut également expliquer ce manquement. En effet, étudier une détérioration définitive de la QdV nécessite de vérifier que celle-ci est maintenue au cours du temps. Enfin, le recul entre ces recommandations et le début de cette revue en 2014 reste beaucoup trop faible. Il serait intéressant de réitérer ce travail, dans quelques années, en couvrant une plus large période. Cela permettrait d'observer une éventuelle amélioration du suivi des recommandations au cours du temps.

Ce travail a permis de souligner l'hétérogénéité des définitions utilisées pour l'approche du temps jusqu'à détérioration. En effet, la plupart des définitions utilisées n'étaient pas adaptées à la situation thérapeutique. Les différentes recommandations proposées jusqu'à maintenant

sont générales et nécessiteraient d'être approfondies. Il existe donc un réel besoin d'élaborer de solides recommandations autour de cette approche.

Des recherches sont en cours autour de la prise en compte du décès, au sein de la méthode du temps jusqu'à détérioration en situation avancée. Le décès entraîne des données manquantes monotones et aucune réelle recommandation n'a été proposée quant à sa prise en compte (Anota et al., 2013). Dans ce contexte thérapeutique, ignorer le décès peut biaiser les résultats. Le décès peut représenter une censure informative puisque son apparition pourrait être liée à l'événement étudié, en l'occurrence la détérioration de la QdV. Une solution pourrait être d'inclure le décès en événement. Cependant, cette approche peut ne pas être optimale dans le cas de décès tardif (i.e. lorsque l'intervalle de temps entre la dernière évaluation de la QdV disponible et la date de décès est trop grand) car elle surestimerait le temps jusqu'à détérioration. Elle refléterait davantage la survie globale si la plupart des événements sont dus au décès. En effet, généralement la QdV n'est plus évaluée depuis un certain temps avant l'occurrence du décès, il sera donc plus facile d'observer comme événement le décès que la détérioration. Il paraît ainsi justifié d'inclure le décès comme un événement lorsqu'il survient dans un délai raisonnable après la dernière évaluation de QdV (Ajani et al., 2007). Une autre solution serait de considérer le décès comme un événement compétitif à notre événement d'intérêt (Shiroiwa et al., 2017). Deux évènements sont dits compétitifs lorsque la survenue de l'un affecte la probabilité d'observer l'autre. Les événements décès et détérioration de la QdV peuvent être considérés comme deux événements compétitifs dits dépendants et exclusifs. En effet, l'occurrence du décès empêchera définitivement d'observer une détérioration de la QdV. Nous avons initié un travail autour de la prise en compte du décès au sein de la méthode du temps jusqu'à détérioration sur plusieurs essais cliniques randomisés en situation avancée ou métastatique. L'objectif de ce travail est de comparer différentes méthodes de prise en compte du décès dans l'approche du temps jusqu'à détérioration. Une étude de simulation pourra être réalisée afin de compléter les résultats. Ce travail permettra de proposer des recommandations adaptées aux décès précoces et tardifs.

Un deuxième objectif de ce travail a été de développer deux macros-programmes SAS pour estimer le temps jusqu'à détérioration. L'ensemble des définitions de détérioration recommandées à ce jour ont été implémentées (Anota et al., 2013; Coens et al., 2020). Ces macros-programmes offrent plusieurs options, comme le choix du score de référence, le choix

de la DMCI, la gestion des données manquantes à l'inclusion et la prise en compte ou non du décès dans la définition de l'événement. Un des avantages de ces deux programmes est que différentes DMCI peuvent être considérées par score. Ils peuvent considérer une détérioration réversible, transitoire ou définitive. L'implémentation de l'approche du temps jusqu'à détérioration avait également été réalisée préalablement sous les logiciels R et Stata (Anota et al., 2017; Bascoul-Mollevi et al., 2018). De plus, le package R permet aussi de calculer les scores de QdV des questionnaires développés par le groupe EORTC. La commande Stata, quant à elle, est spécifique au questionnaire QLQ-C30. Les deux macros-programmes SAS n'ont pas cette particularité. Ils nécessitent le calcul au préalable des scores afin de garantir leur fonctionnement. Ce choix, qui pourrait être considérer comme une limite, permet cependant de les appliquer à tout questionnaire.

Ces deux macros-programmes SAS fournissent des estimations automatiques et fiables de l'approche du temps jusqu'à détérioration. Ils faciliteront l'application de cette approche auprès de la communauté scientifique et permettront d'aider à la standardisation de l'analyse longitudinale des données de PROs dans les essais cliniques en cancérologie. Ainsi, la méthode du temps jusqu'à détérioration sera de plus en plus utilisée et de manière consensuelle. Les différents programmes pourront être régulièrement mis à jour, à la suite de nouvelles propositions de recommandations sur cette approche.

Un troisième objectif a été d'étudier la méthode du temps jusqu'à détérioration au sein de deux études en tant que modalité d'analyse longitudinale des données de QdV. Ces différents travaux ont conduit à s'intéresser à différents enjeux de l'analyse longitudinale : l'effet « *response shift* » et l'occurrence des données manquantes à l'inclusion.

Dans un premier temps, la méthode du temps jusqu'à détérioration a été étudiée dans une cohorte de patientes atteintes d'un cancer du sein en situation adjuvante, qui ont subi une reconstruction mammaire. Cette étude a conduit à s'intéresser à la première détérioration de la patiente, à la gestion de l'absence de randomisation à l'inclusion ainsi qu'à l'impact potentiel d'un effet « *response shift* ».

La définition de détérioration choisie découle des recommandations initialement proposées en situation adjuvante (Anota et al., 2013). En effet, cette définition semble la plus adaptée

puisque les patientes ont de grandes chances de survie.

L'absence de randomisation a été traitée dans l'optique de rendre comparable les deux bras de traitement. Les essais cliniques randomisés sont considérés comme un « *gold standard* » afin de pouvoir constituer des recommandations sur les traitements et réduire les biais de sélection. A ce jour, seule la randomisation permet d'obtenir des groupes similaires tenant compte de leurs caractéristiques. En général, les études sur la reconstruction mammaire sont observationnelles et la faisabilité de la randomisation reste difficile à établir (Winters et al., 2010; Winters et al., 2015). En effet, les différents types de reconstruction mammaire proposés dépendent souvent des caractéristiques et/ou des préférences des patientes mais également de l'expérience du chirurgien (Dauplat et al., 2017; Lee et al., 2009; Winters et al., 2010; Winters et al., 2014). Ainsi, la méthode du score de propension a été utilisée afin de minimiser le biais lié à l'absence de randomisation. Dans cette étude, le score de propension représentait la probabilité qu'une patiente reçoive une reconstruction mammaire spécifique en fonction de ses caractéristiques initialement observées. Ses caractéristiques représentent des facteurs de confusion (Gayat & Porcher, 2012). Les patientes subissant une reconstruction mammaire ALD et LDI, qui ont le même score de propension, peuvent être considérées comme ayant été affectées à l'un des deux groupes par tirage au sort. Le score de propension est couramment estimé par un modèle logistique. Il existe différentes méthodes qui utilisent le score de propension : l'appariement, la stratification, l'ajustement et la pondération par l'inverse du score de propension (« *inverse probability of treatment weighted* »). Nous avons donc choisi d'appliquer la pondération sur le score de propension avec l'idée d'affecter un poids à chaque patiente. L'avantage de cette méthode est de conserver l'ensemble de l'échantillon initial dans l'analyse. La limite est d'obtenir un équilibre entre les groupes uniquement pour les caractéristiques observées contrairement aux essais cliniques randomisés, qui tiennent compte de l'ensemble des caractéristiques observées et non observées.

Des recommandations autour de la méthode du temps jusqu'à détérioration en présence d'un effet « *response shift* » ont également été proposées (Anota et al., 2013). Ces recommandations ont suggéré d'utiliser le meilleur score déjà observé (« *best previous score* ») ou le score précédant la mesure actuelle du patient (« *immediately preceding score* ») comme référence. Cet effet « *response shift* » a déjà été démontré chez les patientes atteintes d'un cancer du sein, en utilisant la méthode « *then-test* » (Dabakuyo et al., 2013). Pour cela, nous avons décidé d'étudier dans cette cohorte l'impact potentiel de cet effet en utilisant le

meilleur score déjà observé comme score de référence. Les résultats n'ont pas pu mettre en évidence un effet « *response shift* » entre les deux types de reconstruction. De plus amples investigations de ces nouveaux scores de référence sont nécessaires pour valider leur application. Ainsi, ces nouveaux scores de référence pourraient être utilisés comme analyse de sensibilité, au sein du temps jusqu'à détérioration, pour caractériser un processus d'adaptation des patients à la maladie et au traitement.

Dans un deuxième temps, la méthode du temps jusqu'à détérioration a été appliquée dans un essai de phase II randomisé chez des patients atteints d'un cancer du pancréas métastatique. Dans cette étude, nous nous sommes intéressés à une détérioration définitive mais également à l'impact de l'occurrence des données manquantes à l'inclusion.

Dans cette étude, deux définitions ont été explorées tenant compte des recommandations initialement proposées en situation avancée ou métastatique. L'analyse principale a été conduite en considérant la survie sans détérioration de la QdV, définie comme une détérioration définitive incluant le décès comme événement. Une analyse de sensibilité a été réalisée en excluant le décès.

La principale limite de cette étude était le pourcentage relativement faible de questionnaires de QdV disponibles à l'inclusion (56,1%). Cependant, ce pourcentage était similaire dans les deux bras de traitement. Une explication probable pourrait être que les deux premiers temps d'évaluation de la QdV se sont déroulés à des intervalles très courts : au moment de la randomisation et au premier jour de traitement. Ceci pourrait avoir causé une certaine confusion au moment de la collecte des données. Les données manquantes à l'inclusion ont un impact direct sur la population d'analyse. Ainsi, une population en intention de traiter modifiée a été définie. Elle correspondait à tous les patients en intention de traiter qui ont au moins un score de QdV disponible à l'inclusion. Cependant, il est recommandé de réaliser les analyses sur la population en intention de traiter, afin de refléter au mieux l'effet du traitement sans induire de biais (Bonnetain et al., 2016). Il est donc essentiel de vérifier que la population en intention de traiter modifiée est représentative de l'ensemble de la population. Cet enjeu a été abordé en analyse de sensibilité, comme le recommande Fairclough (1998) lorsque le pourcentage de données manquantes est entre 20% et 50%. Le profil des données manquantes a été exploré au préalable et n'a été traité qu'à l'inclusion. La détermination du mécanisme des données manquantes à l'inclusion est nécessaire. En effet, la présence du score de QdV à

l'inclusion est primordiale afin de garantir la comparabilité entre les bras de traitement. Elle permet également d'étudier l'évolution du niveau de QdV du patient par rapport à l'inclusion. Dans cette étude, l'hypothèse était que les données manquantes dépendaient des caractéristiques recueillies à l'inclusion. Ainsi, des imputations multiples des items manquants à l'inclusion ont été effectuées selon la méthode de Monte-Carlo par chaînes de Markov, en tenant compte des variables associées aux données manquantes. Par la suite, les analyses du temps jusqu'à détérioration ont été répétées après imputation des données manquantes à l'inclusion. Les différentes analyses ont démontré que ces données manquantes ne semblaient pas biaiser les résultats. Nous avons donc pu conclure que la population en intention de traiter modifiée reflétait bien l'ensemble de la population.

Une perspective de ce travail serait de s'intéresser à la problématique des données manquantes au cours du temps. Dans l'approche du temps jusqu'à détérioration, les données manquantes au cours du temps sont généralement ignorées, considérant que le niveau de QdV du patient reste constant depuis la dernière mesure disponible. Ces données manquantes peuvent néanmoins être informatives du niveau de QdV du patient et ainsi représenter une censure informative. Une solution pourrait être de considérer la donnée manquante comme événement. La méthode « *inverse probability of censoring weighted* », souvent utilisé pour tenir compte des cross-over dans les études randomisées, paraît également intéressante dans ce contexte. Un projet de recherche a été initié afin d'explorer et de comparer ces différentes approches.

La détermination du mécanisme des données manquantes ainsi que les approches statistiques pour les traiter sont d'une importance fondamentale. Dans les essais cliniques, l'enjeu des données manquantes reste rarement étudié (Fiteni et al., 2016). La QdV, étant devenue un critère de jugement important dans les essais cliniques en cancérologie, devrait être évaluée aussi rigoureusement que d'autres critères de jugement classiques afin d'éviter au maximum les données manquantes.

Les différents travaux réalisés ont confirmé le besoin de consolider les recommandations autour de la méthode du temps jusqu'à détérioration. Le développement de consensus similaire à l'idée du projet DATECAN (Bellera et al., 2013) pour le temps jusqu'à détérioration pourrait constituer une avancée vers une méthodologie universelle. En parallèle, la création d'une checklist spécifique à la méthode du temps jusqu'à détérioration devrait être

développée dans le but d'améliorer la qualité de la méthodologie reportée dans les essais. De la même façon, ces checklists devraient être élargies à l'ensemble des méthodes utilisées pour l'analyse longitudinale. La QdV pourra alors être utilisée de plus en plus comme un critère de jugement principal dans les essais cliniques en cancérologie et ainsi devenir un outil d'aide à la décision de la prise en charge des patients. S'inscrivant dans cette trajectoire, le projet en cours Q-RECIST, mené en partenariat avec le groupe EORTC, a pour but de développer et de valider un critère standardisé de type RECIST pour catégoriser les patients en fonction de l'évolution de leur QdV au cours du temps (e.g. amélioration, stabilité et détérioration). En se basant sur le questionnaire QLQ-C30, ce projet se découpe en plusieurs étapes. La première étape est de réaliser une revue de la littérature sur la méthodologie de l'analyse longitudinale des données de QdV dans les essais cliniques randomisés en cancérologie. Cette revue porte sur 4 localisations cancéreuses importantes : sein, colorectal/anal, poumon et prostate. La deuxième étape sera d'élaborer un questionnaire afin de déterminer les éléments pertinents à la construction du critère Q-RECIST. Il sera proposé à un groupe d'experts international afin d'aboutir à un consensus. Il est évident qu'un lien fort existe entre ce critère et la méthode du temps jusqu'à détérioration. Le Q-RECIST devra donc être en adéquation avec les définitions de détérioration qui auront fait l'objet d'une validation collégiale. Une fois validé, le Q-RECIST simplifiera à la fois l'interprétation des résultats et la conception des futurs essais incluant la QdV comme critère principal.

Ces travaux soulignent la nécessité pour la communauté scientifique à redoubler d'effort pour tendre vers une standardisation du temps jusqu'à détérioration et plus largement de la méthodologie de la QdV.

Références

Aaronson, N. K., Ahmedzai, S., Bergman, B., Bullinger, M., Cull, A., Duez, N. J., et al. (1993). 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, 85*(5), 365-376. doi: 10.1093/jnci/85.5.365

Addington-Hall, J., Kalra, L. (2001). Who should measure quality of life? *BMJ, 322*(7299), 1417-1420. doi: 10.1136/bmj.322.7299.1417

Ajani, J. A., Moiseyenko, V. M., Tjulandin, S., Majlis, A., Constenla, M., Boni, C., et al. (2007). Quality of life with docetaxel plus cisplatin and fluorouracil compared with cisplatin and fluorouracil from a phase III trial for advanced gastric or gastroesophageal adenocarcinoma: the V-325 Study Group. *J Clin Oncol, 25*(22), 3210-3216. doi: 10.1200/JCO.2006.08.3956

Anota, A. (2016). Longitudinal analysis of the health-related quality of life in oncology integrating the occurrence of the response shift effect. *Statistica Applicata*.

Anota, A., Boulin, M., Dabakuyo-Yonli, S., Hillon, P., Cercueil, J. P., Minello, A., et al. (2016). An explorative study to assess the association between health-related quality of life and the recommended phase II dose in a phase I trial: idarubicin-loaded beads for chemoembolisation of hepatocellular carcinoma. *BMJ Open, 6*(6), e010696. doi: 10.1136/bmjopen-2015-010696

Anota, A., Hamidou, Z., Paget-Bailly, S., Chibaudel, B., Bascoul-Mollevi, C., Auquier, P., et al. (2013). 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*(1), 5-18. doi: 10.1007/s11136-013-0583-6

Anota, A., Savina, M., Bascoul-Mollevi, C., Bonnetain, F. (2017). Qolr: An r package for the longitudinal analysis of health-related quality of life in oncology. *Journal of Statistical Software, Articles, 77*(12), 1-30.

Arraras, J. I., Greimel, E., Sezer, O., Chie, W. C., Bergenmar, M., Costantini, A., et al. (2010). An international validation study of the EORTC QLQ-INFO25 questionnaire: an

instrument to assess the information given to cancer patients. *Eur J Cancer, 46*(15), 2726-2738. doi: 10.1016/j.ejca.2010.06.118

Basch, E., Deal, A. M., Dueck, A. C., Scher, H. I., Kris, M. G., Hudis, C., et al. (2017). Overall Survival Results of a Trial Assessing Patient-Reported Outcomes for Symptom Monitoring During Routine Cancer Treatment. *JAMA, 318*(2), 197-198. doi: 10.1001/jama.2017.7156

Basch, E., Reeve, B. B., Mitchell, S. A., Clauser, S. B., Minasian, L. M., Dueck, A. C., et al. (2014). Development of the National Cancer Institute's patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE). *J Natl Cancer Inst, 106*(9). doi: 10.1093/jnci/dju244

Bascoul-Mollevi, C., Savina, M., Anota, A., Barbieri, A., Azria, D., Bonnetain, F., et al. (2018). Longitudinal health-related quality of life analysis in oncology with time to event approaches, the STATA command qlqc30_TTD. *Comput Methods Programs Biomed, 158*, 153-159. doi: 10.1016/j.cmpb.2018.02.010

Basen-Engquist, K., Bodurka-Bevers, D., Fitzgerald, M. A., Webster, K., Cella, D., Hu, S., et al. (2001). Reliability and validity of the functional assessment of cancer therapy-ovarian. *J Clin Oncol, 19*(6), 1809-1817. doi: 10.1200/JCO.2001.19.6.1809

Beitz, J., Gnecco, C., Justice, R. (1996). Quality-of-life end points in cancer clinical trials: the U.S. Food and Drug Administration perspective. *J Natl Cancer Inst Monogr*(20), 7-9.

Bellera, C. A., Penel, N., Ouali, M., Bonvalot, S., Casali, P. G., Nielsen, O. S., et al. (2015). Guidelines for time-to-event end point definitions in sarcomas and gastrointestinal stromal tumors (GIST) trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann Oncol, 26*(5), 865-872. doi: 10.1093/annonc/mdu360

Bellera, C. A., Pulido, M., Gourgou, S., Collette, L., Doussau, A., Kramar, A., et al. (2013). Protocol of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project: formal consensus method for the development of guidelines for standardised time-to-event endpoints' definitions in cancer clinical trials. *Eur J Cancer, 49*(4), 769-781. doi: 10.1016/j.ejca.2012.09.035

Bergman, B., Aaronson, N. K., Ahmedzai, S., Kaasa, S., Sullivan, M. (1994). The EORTC

QLQ-LC13: a modular supplement to the EORTC Core Quality of Life Questionnaire (QLQ-C30) for use in lung cancer clinical trials. EORTC Study Group on Quality of Life. *Eur J Cancer, 30A*(5), 635-642. doi: 10.1016/0959-8049(94)90535-5

Bergner, M., Bobbitt, R. A., Pollard, W. E., Martin, D. P., Gilson, B. S. (1976). The sickness impact profile: validation of a health status measure. *Med Care, 14*(1), 57-67.

Bjordal, K., Hammerlid, E., Ahlner-Elmqvist, M., de Graeff, A., Boysen, M., Evensen, J. F., et al. (1999). Quality of life in head and neck cancer patients: validation of the European Organization for Research and Treatment of Cancer Quality of Life Questionnaire-H&N35. *J Clin Oncol, 17*(3), 1008-1019. doi: 10.1200/JCO.1999.17.3.1008

Blanchin, M., Hardouin, J. B., Le Neel, T., Kubis, G., Blanchard, C., Mirallie, E., et al. (2011). Comparison of CTT and Rasch-based approaches for the analysis of longitudinal Patient Reported Outcomes. *Stat Med, 30*(8), 825-838. doi: 10.1002/sim.4153

Blazeby, J. M., Conroy, T., Bottomley, A., Vickery, C., Arraras, J., Sezer, O., et al. (2004). Clinical and psychometric validation of a questionnaire module, the EORTC QLQ-STO 22, to assess quality of life in patients with gastric cancer. *Eur J Cancer, 40*(15), 2260-2268. doi: 10.1016/j.ejca.2004.05.023

Blazeby, J. M., Conroy, T., Hammerlid, E., Fayers, P., Sezer, O., Koller, M., et al. (2003). Clinical and psychometric validation of an EORTC questionnaire module, the EORTC QLQ-OES18, to assess quality of life in patients with oesophageal cancer. *Eur J Cancer, 39*(10), 1384-1394. doi: 10.1016/s0959-8049(03)00270-3

Blazeby, J. M., Fayers, P., Conroy, T., Sezer, O., Ramage, J., Rees, M., et al. (2009). Validation of the European Organization for Research and Treatment of Cancer QLQ-LMC21 questionnaire for assessment of patient-reported outcomes during treatment of colorectal liver metastases. *Br J Surg, 96*(3), 291-298. doi: 10.1002/bjs.6471

Bonnetain, F. (2010). [Health related quality of life and endpoints in oncology]. *Cancer Radiother, 14*(6-7), 515-518. doi: 10.1016/j.canrad.2010.06.001

Bonnetain, F., Bonsing, B., Conroy, T., Dousseau, A., Glimelius, B., Haustermans, K., et al. (2014). Guidelines for time-to-event end-point definitions in trials for pancreatic cancer. Results of the DATECAN initiative (Definition for the Assessment of Time-to-event End-points in CANcer trials). *Eur J Cancer, 50*(17), 2983-2993. doi: 10.1016/j.ejca.2014.07.011

Bonnetain, F., Borg, C., Adams, R. R., Ajani, J. A., Benson, A., Bleiberg, H., et al. (2017). How health-related quality of life assessment should be used in advanced colorectal cancer clinical trials. *Ann Oncol, 28*(9), 2077-2085. doi: 10.1093/annonc/mdx191

Bonnetain, F., Dahan, L., Maillard, E., Ychou, M., Mitry, E., Hammel, P., et al. (2010). 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*(15), 2753-2762. doi: 10.1016/j.ejca.2010.07.023

Bonnetain, F., Fiteni, F., Efficace, F., Anota, A. (2016). Statistical Challenges in the Analysis of Health-Related Quality of Life in Cancer Clinical Trials. *J Clin Oncol, 34*(16), 1953-1956. doi: 10.1200/JCO.2014.56.7974

Bottomley, A., Pe, M., Sloan, J., Basch, E., Bonnetain, F., Calvert, M., et al. (2016). Analysing data from patient-reported outcome and quality of life endpoints for cancer clinical trials: a start in setting international standards. *Lancet Oncol, 17*(11), e510-e514. doi: 10.1016/S1470-2045(16)30510-1

Boye, M., Wang, X., Srimuninnimit, V., Kang, J. H., Tsai, C. M., Orlando, M., et al. (2016). First-Line Pemetrexed Plus Cisplatin Followed by Gefitinib Maintenance Therapy Versus Gefitinib Monotherapy in East Asian Never-Smoker Patients With Locally Advanced or Metastatic Nonsquamous Non-Small-cell Lung Cancer: Quality of Life Results From a Randomized Phase III Trial. *Clin Lung Cancer, 17*(2), 150-160. doi: 10.1016/j.cllc.2015.12.004

Brady, M. J., Cella, D. F., Mo, F., Bonomi, A. E., Tulsky, D. S., Lloyd, S. R., et al. (1997). Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. *J Clin Oncol, 15*(3), 974-986. doi: 10.1200/JCO.1997.15.3.974

Brahmer, J. R., Rodríguez-Abreu, D., Robinson, A. G., Hui, R., Csőszi, T., Fülöp, A., et al. (2017). 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*(12), 1600-1609. doi: 10.1016/S1470-2045(17)30690-3

Brandes, A. A., Finocchiaro, G., Zagonel, V., Reni, M., Caserta, C., Fabi, A., et al. (2016). AVAREG: a phase II, randomized, noncomparative study of fotemustine or bevacizumab for patients with recurrent glioblastoma. *Neuro Oncol, 18*(9), 1304-1312. doi:

10.1093/neuonc/now035

Bucquet, D., Condon, S., Ritchie, K. (1990). The French version of the Nottingham Health Profile. A comparison of items weights with those of the source version. *Soc Sci Med, 30*(7), 829-835. doi: 10.1016/0277-9536(90)90207-9

Calvert, M., Blazeby, J., Altman, D. G., Revicki, D. A., Moher, D., Brundage, M. D., et al. (2013). Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA, 309*(8), 814-822. doi: 10.1001/jama.2013.879

Calvert, M., Kyte, D., Mercieca-Bebber, R., Slade, A., Chan, A. W., King, M. T., et al. (2018). Guidelines for Inclusion of Patient-Reported Outcomes in Clinical Trial Protocols: The SPIRIT-PRO Extension. *JAMA, 319*(5), 483-494. doi: 10.1001/jama.2017.21903

Carr, A. J., Gibson, B., Robinson, P. G. (2001). Measuring quality of life: Is quality of life determined by expectations or experience? *BMJ, 322*(7296), 1240-1243. doi: 10.1136/bmj.322.7296.1240

Carr, A. J., Higginson, I. J. (2001). Are quality of life measures patient centred? *BMJ, 322*(7298), 1357-1360. doi: 10.1136/bmj.322.7298.1357

Cella, D. F., Bonomi, A. E., Lloyd, S. R., Tulsky, D. S., Kaplan, E., Bonomi, P. (1995). Reliability and validity of the Functional Assessment of Cancer Therapy-Lung (FACT-L) quality of life instrument. *Lung Cancer, 12*(3), 199-220. doi: 10.1016/0169-5002(95)00450-f

Cella, D. F., Jensen, S. E., Webster, K., Hongyan, D., Lai, J. S., Rosen, S., et al. (2012). Measuring health-related quality of life in leukemia: the Functional Assessment of Cancer Therapy--Leukemia (FACT-Leu) questionnaire. *Value Health, 15*(8), 1051-1058. doi: 10.1016/j.jval.2012.08.2210

Cella, D. F., Tulsky, D. S., Gray, G., Sarafian, B., Linn, E., Bonomi, A., et al. (1993). The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol, 11*(3), 570-579. doi: 10.1200/JCO.1993.11.3.570

Chie, W. C., Blazeby, J. M., Hsiao, C. F., Chiu, H. C., Poon, R. T., Mikoshiba, N., et al. (2012). International cross-cultural field validation of an European Organization for Research and Treatment of Cancer questionnaire module for patients with primary liver cancer, the European Organization for Research and Treatment of Cancer quality-of-life questionnaire HCC18. *Hepatology, 55*(4), 1122-1129. doi: 10.1002/hep.24798

Chow, E., Nguyen, J., Zhang, L., Tseng, L. M., Hou, M. F., Fairchild, A., et al. (2012). International field testing of the reliability and validity of the EORTC QLQ-BM22 module to assess health-related quality of life in patients with bone metastases. *Cancer, 118*(5), 1457-1465. doi: 10.1002/cncr.26410

Chwalow, A. J., Lurie, A., Bean, K., Parent du Chatelet, I., Venot, A., Dusser, D., et al. (1992). A French version of the Sickness Impact Profile (SIP): stages in the cross cultural validation of a generic quality of life scale. *Fundam Clin Pharmacol, 6*(7), 319-326. doi: 10.1111/j.1472-8206.1992.tb00126.x

Cnaan, A., Laird, N. M., Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med, 16*(20), 2349-2380.

Cocks, K., Cohen, D., Wisloff, F., Sezer, O., Lee, S., Hippe, E., et al. (2007). An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-MY20) in assessing the quality of life of patients with multiple myeloma. *Eur J Cancer, 43*(11), 1670-1678. doi: 10.1016/j.ejca.2007.04.022

Cocks, K., King, M. T., Velikova, G., de Castro, G., Jr., Martyn St-James, M., Fayers, P. M., et al. (2012). Evidence-based guidelines for interpreting change scores for the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *Eur J Cancer, 48*(11), 1713-1721. doi: 10.1016/j.ejca.2012.02.059

Coens, C., Pe, M., Dueck, A. C., Sloan, J., Basch, E., Calvert, M., et al. (2020). International standards for the analysis of quality-of-life and patient-reported outcome endpoints in cancer randomised controlled trials: recommendations of the SISAQOL Consortium. *Lancet Oncol, 21*(2), e83-e96. doi: 10.1016/S1470-2045(19)30790-9

Cohen, R., Vernerey, D., Bellera, C., Meurisse, A., Henriques, J., Paoletti, X., et al. (2020). Guidelines for time-to-event end-point definitions in adjuvant randomised trials for patients with localised colon cancer: Results of the DATECAN initiative. *Eur J Cancer, 130*, 63-71. doi: 10.1016/j.ejca.2020.02.009

Conde-Sala, J. L., Garre-Olmo, J., Turro-Garriga, O., Lopez-Pousa, S., Vilalta-Franch, J. (2009). Factors related to perceived quality of life in patients with Alzheimer's disease: the patient's perception compared with that of caregivers. *Int J Geriatr Psychiatry, 24*(6), 585-594. doi: 10.1002/gps.2161

Conroy, T., Brédart, A., Briançon, S. (2005). Valeur pronostique des mesures de qualité de vie en cancérologie. *Revue francophone de psycho-oncologie, 4*(1), 23-28.

Conroy, T., Mercier, M., Bonneterre, J., Luporsi, E., Lefebvre, J. L., Lapeyre, M., et al. (2004). French version of FACT-G: validation and comparison with other cancer-specific instruments. *Eur J Cancer, 40*(15), 2243-2252. doi: 10.1016/j.ejca.2004.06.010

Copay, A. G., Subach, B. R., Glassman, S. D., Polly, D. W., Jr., Schuler, T. C. (2007). Understanding the minimum clinically important difference: a review of concepts and methods. *Spine J, 7*(5), 541-546. doi: 10.1016/j.spinee.2007.01.008

Cormier, J. N., Ross, M. I., Gershenwald, J. E., Lee, J. E., Mansfield, P. F., Camacho, L. H., et al. (2008). Prospective assessment of the reliability, validity, and sensitivity to change of the Functional Assessment of Cancer Therapy-Melanoma questionnaire. *Cancer, 112*(10), 2249-2257. doi: 10.1002/cncr.23424

Cortes, J., Hudgens, S., Twelves, C., Perez, E. A., Awada, A., Yelle, L., et al. (2015). Health-related quality of life in patients with locally advanced or metastatic breast cancer treated with eribulin mesylate or capecitabine in an open-label randomized phase 3 trial. *Breast Cancer Res Treat, 154*(3), 509-520. doi: 10.1007/s10549-015-3633-7

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187-202.

Crosby, R. D., Kolotkin, R. L., Williams, G. R. (2003). Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol, 56*(5), 395-407. doi: 10.1016/s0895-4356(03)00044-1

Dabakuyo, T. S., Guillemin, F., Conroy, T., Velten, M., Jolly, D., Mercier, M., et al. (2013). Response shift effects on measuring post-operative quality of life among breast cancer patients: a multicenter cohort study. *Qual Life Res, 22*(1), 1-11. doi: 10.1007/s11136-012-0135-5

Darling, G., Eton, D. T., Sulman, J., Casson, A. G., Celia, D. (2006). Validation of the functional assessment of cancer therapy esophageal cancer subscale. *Cancer, 107*(4), 854-863. doi: 10.1002/cncr.22055

Dauplat, J., Kwiatkowski, F., Rouanet, P., Delay, E., Clough, K., Verhaeghe, J. L., et al. (2017). Quality of life after mastectomy with or without immediate breast reconstruction. *Br J*

Surg, 104(9), 1197-1206. doi: 10.1002/bjs.10537

Degboe, A., Ivanescu, C., Rohay, J. M., Turner, R. R., Cella, D. (2019). Validity and performance of the Functional Assessment of Cancer Therapy-Bladder (FACT-Bl) among advanced urothelial cancer patients. *Support Care Cancer, 27*(11), 4189-4198. doi: 10.1007/s00520-019-04709-0

Detmar, S. B., Muller, M. J., Schornagel, J. H., Wever, L. D., Aaronson, N. K. (2002). Health-related quality-of-life assessments and patient-physician communication: a randomized controlled trial. *JAMA, 288*(23), 3027-3034. doi: 10.1001/jama.288.23.3027

Doward, L. C., McKenna, S. P. (2004). Defining patient-reported outcomes. *Value Health, 7 Suppl 1*, S4-8. doi: 10.1111/j.1524-4733.2004.7s102.x

Ediebah, D. E., Quinten, C., Coens, C., Ringash, J., Dancey, J., Zikos, E., et al. (2018). Quality of life as a prognostic indicator of survival: A pooled analysis of individual patient data from canadian cancer trials group clinical trials. *Cancer, 124*(16), 3409-3416. doi: 10.1002/cncr.31556

EORTC Quality of Life Group (April 2011). Guidelines for Developing Questionnaire Modules. Fourth Edition.

Esper, P., Mo, F., Chodak, G., Sinner, M., Cella, D., Pienta, K. J. (1997). Measuring quality of life in men with prostate cancer using the functional assessment of cancer therapy-prostate instrument. *Urology, 50*(6), 920-928. doi: 10.1016/S0090-4295(97)00459-7

Fairclough, D. L. (1998). Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Stat Med, 17*(5-7), 781-796. doi: 10.1002/(sici)1097-0258(19980315/15)17:5/7<781::aid-sim821>3.0.co;2-o

Fairclough, D. L. (2010). *Design and analysis of quality of life studies in clinical trials*: Chapman and Hall/CRC.

Fairclough, D. L., Peterson, H. F., Chang, V. (1998). Why are missing quality of life data a problem in clinical trials of cancer therapy? *Stat Med, 17*(5-7), 667-677. doi: 10.1002/(sici)1097-0258(19980315/15)17:5/7<667::aid-sim813>3.0.co;2-6

Fayers, P. M., Aaronson, N. K., Bjordal, K., Groenvold, M., Curran, D., Bottomley, A., et al. (2001). The EORTC QLQ-C30 Scoring Manual (3rd edition).

Fayers, P. M., Jones, D. R. (1983). Measuring and analysing quality of life in cancer clinical trials: a review. *Stat Med, 2*(4), 429-446. doi: 10.1002/sim.4780020402

Fayers, P. M., Machin, D. (2000). *Quality of Life : Assessment, Analysis and Interpretation*: Wiley.

FDA (2006). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes, 4*, 79. doi: 10.1186/1477-7525-4-79

FDA (2009). Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims.

FDA (2019). Clinical trial endpoints for the approval of cancer drugs and biologics guidance for industry.

Feld, R. (1995). Endpoints in cancer clinical trials: is there a need for measuring quality of life? *Support Care Cancer, 3*(1), 23-27. doi: 10.1007/bf00343917

Fielding, S. (2008). Simple imputation methods were inadequate for missing not at random (MNAR) quality of life data. *Health Qual Life Outcomes, 6*, 57. doi: 10.1186/1477-7525-6-57

Fielding, S., Fayers, P., Ramsay, C. R. (2012). Analysing randomised controlled trials with missing data: choice of approach affects conclusions. *Contemp Clin Trials, 33*(3), 461-469. doi: 10.1016/j.cct.2011.12.002

Fielding, S., Maclennan, G., Cook, J. A., Ramsay, C. R. (2008). A review of RCTs in four medical journals to assess the use of imputation to overcome missing data in quality of life outcomes. *Trials, 9*, 51. doi: 10.1186/1745-6215-9-51

Fielding, S., Ogbuagu, A., Sivasubramaniam, S., MacLennan, G., Ramsay, C. R. (2016). Reporting and dealing with missing quality of life data in RCTs: has the picture changed in the last decade? *Qual Life Res, 25*(12), 2977-2983. doi: 10.1007/s11136-016-1411-6

Fiteni, F., Anota, A., Westeel, V., Bonnetain, F. (2015). [Health-related quality of life in phase III cancer clinical trials: from questionnaire administration to statistical analysis]. *Bull Cancer, 102*(4), 360-366. doi: 10.1016/j.bulcan.2015.02.014

Fiteni, F., Anota, A., Westeel, V., Bonnetain, F. (2016). Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a

critical review. *BMC Cancer, 16*, 122. doi: 10.1186/s12885-016-2152-1

Fiteni, F., Ray, I. L., Ousmen, A., Isambert, N., Anota, A., Bonnetain, F. (2019). Health-related quality of life as an endpoint in oncology phase I trials: a systematic review. *BMC Cancer, 19*(1), 361. doi: 10.1186/s12885-019-5579-3

Fiteni, F., Westeel, V., Pivot, X., Borg, C., Vernerey, D., Bonnetain, F. (2014). Endpoints in cancer clinical trials. *J Visc Surg, 151*(1), 17-22. doi: 10.1016/j.jviscsurg.2013.10.001

Freemantle, N., Calvert, M., Wood, J., Eastaugh, J., Griffin, C. (2003). Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA, 289*(19), 2554-2559. doi: 10.1001/jama.289.19.2554

Gayat, E., Porcher, R. (2012). Comparaison de l'efficacité de deux thérapeutiques en l'absence de randomisation: intérêts et limites des méthodes utilisant les scores de propension. *Réanimation, 21*(1), 109-116.

Gill, T. M., Feinstein, A. R. (1994). A critical appraisal of the quality of quality-of-life measurements. *JAMA, 272*(8), 619-626.

Gnanasakthy, A., Barrett, A., Evans, E., D'Alessio, D., Romano, C. D. (2019). A Review of Patient-Reported Outcomes Labeling for Oncology Drugs Approved by the FDA and the EMA (2012-2016). *Value Health, 22*(2), 203-209. doi: 10.1016/j.jval.2018.09.2842

Goel, M. K., Khanna, P., Kishore, J. (2010). Understanding survival analysis: Kaplan-Meier estimate. *Int J Ayurveda Res, 1*(4), 274-278. doi: 10.4103/0974-7788.76794

Gourgou-Bourgade, S., Cameron, D., Poortmans, P., Asselain, B., Azria, D., Cardoso, F., et al. (2015). Guidelines for time-to-event end point definitions in breast cancer trials: results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann Oncol, 26*(12), 2505-2506. doi: 10.1093/annonc/mdv478

Greimel, E., Bottomley, A., Cull, A., Waldenstrom, A. C., Arraras, J., Chauvenet, L., et al. (2003). An international field study of the reliability and validity of a disease-specific questionnaire module (the QLQ-OV28) in assessing the quality of life of patients with ovarian cancer. *Eur J Cancer, 39*(10), 1402-1408. doi: 10.1016/s0959-8049(03)00307-1

Greimel, E., Kuljanic Vlasic, K., Waldenstrom, A. C., Duric, V. M., Jensen, P. T., Singer, S., et al. (2006). The European Organization for Research and Treatment of Cancer (EORTC)

Quality-of-Life questionnaire cervical cancer module: EORTC QLQ-CX24. *Cancer, 107*(8), 1812-1822. doi: 10.1002/cncr.22217

Greimel, E., Nordin, A., Lanceley, A., Creutzberg, C. L., van de Poll-Franse, L. V., Radisic, V. B., et al. (2011). Psychometric validation of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire-Endometrial Cancer Module (EORTC QLQ-EN24). *Eur J Cancer, 47*(2), 183-190. doi: 10.1016/j.ejca.2010.08.014

Groenvold, M., Petersen, M. A., Aaronson, N. K., Arraras, J. I., Blazeby, J. M., Bottomley, A., et al. (2006). The development of the EORTC QLQ-C15-PAL: a shortened questionnaire for cancer patients in palliative care. *Eur J Cancer, 42*(1), 55-64. doi: 10.1016/j.ejca.2005.06.022

Gupta, S. K. (2011). Intention-to-treat concept: A review. *Perspect Clin Res, 2*(3), 109-112. doi: 10.4103/2229-3485.83221

Halkett, G. K. B., Wigley, C. A., Aoun, S. M., Portaluri, M., Tramacere, F., Livi, L., et al. (2018). International validation of the EORTC QLQ-PRT20 module for assessment of quality of life symptoms relating to radiation proctitis: a phase IV study. *Radiat Oncol, 13*(1), 162. doi: 10.1186/s13014-018-1107-x

Hamel, J. F., Saulnier, P., Pe, M., Zikos, E., Musoro, J., Coens, C., et al. (2017). A systematic review of the quality of statistical methods employed for analysing quality of life data in cancer randomised controlled trials. *Eur J Cancer, 83*, 166-176. doi: 10.1016/j.ejca.2017.06.025

Harbeck, N., Iyer, S., Turner, N., Cristofanilli, M., Ro, J., Andre, F., et al. (2016). Quality of life with palbociclib plus fulvestrant in previously treated hormone receptor-positive, HER2-negative metastatic breast cancer: patient-reported outcomes from the PALOMA-3 trial. *Ann Oncol, 27*(6), 1047-1054. doi: 10.1093/annonc/mdw139

Hardouin, J. B., Conroy, R., Sebille, V. (2011). Imputation by the mean score should be avoided when validating a Patient Reported Outcomes questionnaire by a Rasch model in presence of informative missing data. *BMC Med Res Methodol, 11*, 105. doi: 10.1186/1471-2288-11-105

Heffernan, N., Cella, D., Webster, K., Odom, L., Martone, M., Passik, S., et al. (2002). Measuring health-related quality of life in patients with hepatobiliary cancers: the functional

assessment of cancer therapy-hepatobiliary questionnaire. *J Clin Oncol, 20*(9), 2229-2239. doi: 10.1200/JCO.2002.07.093

Hjermstad, M. J., Bergenmar, M., Bjordal, K., Fisher, S. E., Hofmeister, D., Montel, S., et al. (2016). International field testing of the psychometric properties of an EORTC quality of life module for oral health: the EORTC QLQ-OH15. *Support Care Cancer, 24*(9), 3915-3924. doi: 10.1007/s00520-016-3216-0

Huisman, M. (2000). Imputation of missing item responses: Some simple techniques. *Quality and Quantity, 34*(4), 331-351.

Hunt, S. M., McKenna, S. P., McEwen, J., Backett, E. M., Williams, J., Papp, E. (1980). A quantitative approach to perceived health status: a validation study. *J Epidemiol Community Health, 34*(4), 281-286.

Jaeschke, R., Singer, J., Guyatt, G. H. (1989). Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials, 10*(4), 407-415. doi: 10.1016/0197-2456(89)90005-6

Janda, M., Obermair, A., Cella, D., Perrin, L. C., Nicklin, J. L., Ward, B. G., et al. (2005). The functional assessment of cancer-vulvar: reliability and validity. *Gynecol Oncol, 97*(2), 568-575. doi: 10.1016/j.ygyno.2005.01.047

Johnson, J. R., Temple, R. (1985). Food and Drug Administration requirements for approval of new anticancer drugs. *Cancer Treat Rep, 69*(10), 1155-1159.

Joly, F., Hilpert, F., Okamoto, A., Stuart, G., Ochiai, K., Friedlander, M. (2017). Fifth Ovarian Cancer Consensus Conference of the Gynecologic Cancer InterGroup: Recommendations on incorporating patient-reported outcomes in clinical trials in epithelial ovarian cancer. *Eur J Cancer, 78*, 133-138. doi: 10.1016/j.ejca.2017.03.019

Karimi, M., Brazier, J. (2016). Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? *Pharmacoeconomics, 34*(7), 645-649. doi: 10.1007/s40273-016-0389-9

Karnofsky, D. A. (1949). The clinical evaluation of chemotherapeutic agents in cancer. *Evaluation of chemotherapeutic agents*, 191-205.

Kaupp-Roberts, S. D., Yadegarfar, G., Friend, E., O'Donnell, C. M., Valle, J. W., Byrne, C.,

et al. (2016). Validation of the EORTC QLQ-BIL21 questionnaire for measuring quality of life in patients with cholangiocarcinoma and cancer of the gallbladder. *Br J Cancer, 115*(9), 1032-1038. doi: 10.1038/bjc.2016.284

Kiebert, G. M., Curran, D., Aaronson, N. K. (1998). Quality of life as an endpoint in EORTC clinical trials. European Organization for Research and Treatment for Cancer. *Stat Med, 17*(5-7), 561-569. doi: 10.1002/(sici)1097-0258(19980315/15)17:5/7<561::aid-sim803>3.0.co;2-s

Kramar, A., Negrier, S., Sylvester, R., Joniau, S., Mulders, P., Powles, T., et al. (2015). Guidelines for the definition of time-to-event end points in renal cell cancer clinical trials: results of the DATECAN project. *Ann Oncol, 26*(12), 2392-2398. doi: 10.1093/annonc/mdv380

Lagergren, P., Fayers, P., Conroy, T., Stein, H. J., Sezer, O., Hardwick, R., et al. (2007). Clinical and psychometric validation of a questionnaire module, the EORTC QLQ-OG25, to assess health-related quality of life in patients with cancer of the oesophagus, the oesophago-gastric junction and the stomach. *Eur J Cancer, 43*(14), 2066-2073. doi: 10.1016/j.ejca.2007.07.005

Lee, C., Sunu, C., Pignone, M. (2009). Patient-reported outcomes of breast reconstruction after mastectomy: a systematic review. *J Am Coll Surg, 209*(1), 123-133. doi: 10.1016/j.jamcollsurg.2009.02.061

Leidy, N. K., Revicki, D. A., Geneste, B. (1999). Recommendations for evaluating the validity of quality of life claims for labeling and promotion. *Value Health, 2*(2), 113-127. doi: 10.1046/j.1524-4733.1999.02210.x

Leplège, A., Ecosse, E., Verdier, A., Perneger, T. V. (1998). The French SF-36 Health Survey: translation, cultural adaptation and preliminary psychometric evaluation. *J Clin Epidemiol, 51*(11), 1013-1023. doi: 10.1016/s0895-4356(98)00093-6

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American statistical Association, 83*(404), 1198-1202.

Little, R. J., Rubin, D. B. (1987). *Statistical analysis with missing data*: John Wiley & Sons.

Lydick, E., Epstein, R. S. (1993). Interpretation of quality of life changes. *Qual Life Res, 2*(3), 221-226. doi: 10.1007/BF00435226

Ma Santé 2022 (2018). Ma santé 2022: un engagement collectif.

Maio, M., Scherpereel, A., Calabro, L., Aerts, J., Cedres Perez, S., Bearz, A., et al. (2017). Tremelimumab as second-line or third-line treatment in relapsed malignant mesothelioma (DETERMINE): a multicentre, international, randomised, double-blind, placebo-controlled phase 2b trial. *Lancet Oncol, 18*(9), 1261-1273. doi: 10.1016/S1470-2045(17)30446-1

Mathoulin-Pelissier, S., Gourgou-Bourgade, S., Bonnetain, F., Kramar, A. (2008). Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J Clin Oncol, 26*(22), 3721-3726. doi: 10.1200/JCO.2007.14.1192

Mercier, M., Schraub, S. (2005). Qualité de vie: quels outils de mesure? *27° Journées de la Société française de sénologie et de pathologie mammaire, Deauville*.

Methy, N., Bedenne, L., Bonnetain, F. (2010). Surrogate endpoints for overall survival in digestive oncology trials: which candidates? A questionnaires survey among clinicians and methodologists. *BMC Cancer, 10*, 277. doi: 10.1186/1471-2407-10-277

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res, 19*(4), 539-549. doi: 10.1007/s11136-010-9606-8

Montori, V. M., Permanyer-Miralda, G., Ferreira-Gonzalez, I., Busse, J. W., Pacheco-Huergo, V., Bryant, D., et al. (2005). Validity of composite end points in clinical trials. *BMJ, 330*(7491), 594-596. doi: 10.1136/bmj.330.7491.594

Moons, P., Budts, W., De Geest, S. (2006). Critique on the conceptualisation of quality of life: a review and evaluation of different conceptual approaches. *Int J Nurs Stud, 43*(7), 891-901. doi: 10.1016/j.ijnurstu.2006.03.015

Musoro, Z. J., Hamel, J. F., Ediebah, D. E., Cocks, K., King, M. T., Groenvold, M., et al. (2018). Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol. *BMJ Open, 8*(1), e019117. doi: 10.1136/bmjopen-2017-019117

Oort, F. J. (2005). Using structural equation modeling to detect response shifts and true change. *Qual Life Res, 14*(3), 587-598. doi: 10.1007/s11136-004-0830-y

Osoba, D. (1992). The Quality of Life Committee of the Clinical Trials Group of the National Cancer Institute of Canada: organization and functions. *Qual Life Res, 1*(3), 211-218. doi: 10.1007/bf00635620

Osoba, D. (1994). Lessons learned from measuring health-related quality of life in oncology. *J Clin Oncol, 12*(3), 608-616. doi: 10.1200/JCO.1994.12.3.608

Osoba, D. (2007). Translating the science of patient-reported outcomes assessment into clinical practice. *J Natl Cancer Inst Monogr*(37), 5-11. doi: 10.1093/jncimonographs/lgm002

Osoba, D. (2011). Health-related quality of life and cancer clinical trials. *Ther Adv Med Oncol, 3*(2), 57-71. doi: 10.1177/1758834010395342

Osoba, D., Rodrigues, G., Myles, J., Zee, B., Pater, J. (1998). Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol, 16*(1), 139-144. doi: 10.1200/JCO.1998.16.1.139

Panageas, K. S., Ben-Porat, L., Dickler, M. N., Chapman, P. B., Schrag, D. (2007). When you look matters: the effect of assessment schedule on progression-free survival. *J Natl Cancer Inst, 99*(6), 428-432. doi: 10.1093/jnci/djk091

Pazart, L., Cretin, E., Grodard, G., Cornet, C., Mathieu-Nicot, F., Bonnetain, F., et al. (2014). Parenteral nutrition at the palliative phase of advanced cancer: the ALIM-K study protocol for a randomized controlled trial. *Trials, 15*, 370. doi: 10.1186/1745-6215-15-370

Pe, M., Dorme, L., Coens, C., Basch, E., Calvert, M., Campbell, A., et al. (2018). Statistical analysis of patient-reported outcome data in randomised controlled trials of locally advanced and metastatic breast cancer: a systematic review. *Lancet Oncol, 19*(9), e459-e469. doi: 10.1016/S1470-2045(18)30418-2

Pickard, A. S., Knight, S. J. (2005). Proxy evaluation of health-related quality of life: a conceptual framework for understanding multiple proxy perspectives. *Med Care, 43*(5), 493-499. doi: 10.1097/01.mlr.0000160419.27642.a8

Quinten, C., Martinelli, F., Coens, C., Sprangers, M. A., Ringash, J., Gotay, C., et al. (2014). A global analysis of multitrial data investigating quality of life and symptoms as prognostic factors for survival in different tumor sites. *Cancer, 120*(2), 302-311. doi: 10.1002/cncr.28382

Reck, M., Brahmer, J., Bennett, B., Taylor, F., Penrod, J. R., DeRosa, M., et al. (2018).

Evaluation of health-related quality of life and symptoms in patients with advanced non-squamous non-small cell lung cancer treated with nivolumab or docetaxel in CheckMate 057. *Eur J Cancer, 102*, 23-30. doi: 10.1016/j.ejca.2018.05.005

Revicki, D. A., Gold, K., Buckman, D., Chan, K., Kallich, J. D., Woolley, J. M. (2001). Imputing physical health status scores missing owing to mortality: results of a simulation comparing multiple techniques. *Med Care, 39*(1), 61-71. doi: 10.1097/00005650-200101000-00008

Revicki, D. A., Hays, R. D., Cella, D., Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol, 61*(2), 102-109.

Roydhouse, J. K. (2019). Blinding and Patient-Reported Outcome Completion Rates in US Food and Drug Administration Cancer Trial Submissions, 2007-2017. *J Natl Cancer Inst, 111*(5), 459-464. doi: 10.1093/jnci/djy181

Roydhouse, J. K., Fiero, M. H., Kluetz, P. G. (2019). Investigating Potential Bias in Patient-Reported Outcomes in Open-label Cancer Trials. *JAMA Oncology, 5*(4), 457-458. doi: 10.1001/jamaoncol.2018.6205

Royston, P., Parmar, M. K. (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med, 30*(19), 2409-2421. doi: 10.1002/sim.4274

Royston, P., Parmar, M. K. (2013). Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol, 13*, 152. doi: 10.1186/1471-2288-13-152

Rubin, D. B. (2004). *Multiple imputation for nonresponse in surveys* (Vol. 81): John Wiley & Sons.

Satoh, T., Bang, Y. J., Gotovkin, E. A., Hamamoto, Y., Kang, Y. K., Moiseyenko, V. M., et al. (2014). Quality of life in the trastuzumab for gastric cancer trial. *Oncologist, 19*(7), 712-719. doi: 10.1634/theoncologist.2014-0058

Schadendorf, D., Dummer, R., Hauschild, A., Robert, C., Hamid, O., Daud, A., et al. (2016). Health-related quality of life in the randomised KEYNOTE-002 study of pembrolizumab versus chemotherapy in patients with ipilimumab-refractory melanoma. *Eur J Cancer, 67*, 46-

54. doi: 10.1016/j.ejca.2016.07.018

Schünemann, H. J., Guyatt, G. H. (2005). Commentary—goodbye M (C) ID! Hello MID, where do you come from? *Health services research, 40*(2), 593-597.

Schwartz, C. E., Sprangers, M. A. (1999). Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med, 48*(11), 1531-1548. doi: 10.1016/s0277-9536(99)00047-7

Secord, A. A., Coleman, R. L., Havrilesky, L. J., Abernethy, A. P., Samsa, G. P., Cella, D. (2015). Patient-reported outcomes as end points and outcome indicators in solid tumours. *Nat Rev Clin Oncol, 12*(6), 358-370. doi: 10.1038/nrclinonc.2015.29

Shiroiwa, T., Fukuda, T., Shimozuma, K., Mouri, M., Hagiwara, Y., Doihara, H., et al. (2017). Long-term health status as measured by EQ-5D among patients with metastatic breast cancer: comparison of first-line oral S-1 and taxane therapies in the randomized phase III SELECT BC trial. *Qual Life Res, 26*(2), 445-453. doi: 10.1007/s11136-016-1388-1

Shrestha, A., Martin, C., Burton, M., Walters, S., Collins, K., Wyld, L. (2019). Quality of life versus length of life considerations in cancer patients: A systematic literature review. *Psychooncology, 28*(7), 1367-1380. doi: 10.1002/pon.5054

Singer, S., Araujo, C., Arraras, J. I., Baumann, I., Boehm, A., Brokstad Herlofson, B., et al. (2015). Measuring quality of life in patients with head and neck cancer: Update of the EORTC QLQ-H&N Module, Phase III. *Head Neck, 37*(9), 1358-1367. doi: 10.1002/hed.23762

Sloan, J. A., Dueck, A. C., Erickson, P. A., Guess, H., Revicki, D. A., Santanello, N. C., et al. (2007). Analysis and interpretation of results based on patient-reported outcomes. *Value Health, 10 Suppl 2*, S106-115. doi: 10.1111/j.1524-4733.2007.00273.x

Sodergren, S. C., Johnson, C. D., Gilbert, A., Tomaszewski, K. A., Chu, W., Chung, H. T., et al. (2018). Phase I-III development of the EORTC QLQ-ANL27, a health-related quality of life questionnaire for anal cancer. *Radiother Oncol, 126*(2), 222-228. doi: 10.1016/j.radonc.2017.11.018

Sprangers, M. A., Groenvold, M., Arraras, J. I., Franklin, J., te Velde, A., Muller, M., et al. (1996). The European Organization for Research and Treatment of Cancer breast cancer-specific quality-of-life questionnaire module: first results from a three-country field study. *J*

Clin Oncol, 14(10), 2756-2768. doi: 10.1200/JCO.1996.14.10.2756

Sprangers, M. A., Schwartz, C. E. (1999). Integrating response shift into health-related quality of life research: a theoretical model. *Soc Sci Med, 48*(11), 1507-1515. doi: 10.1016/s0277-9536(99)00045-3

Sprangers, M. A., Van Dam, F. S., Broersen, J., Lodder, L., Wever, L., Visser, M. R., et al. (1999). Revealing response shift in longitudinal research on fatigue--the use of the thentest approach. *Acta Oncol, 38*(6), 709-718.

Sridhara, R., Mandrekar, S. J., Dodd, L. E. (2013). Missing data and measurement variability in assessing progression-free survival endpoint in randomized clinical trials. *Clin Cancer Res, 19*(10), 2613-2620. doi: 10.1158/1078-0432.CCR-12-2938

Sztankay, M., Aaronson, N. K., Arraras, J. I., Basso, U., Bumbasirevic, U., Efficace, F., et al. (2018). International phase IV validation study of an EORTC quality of life questionnaire for testicular cancer patients: the EORTC QLQ-TC26. *BMC Cancer, 18*(1), 1104. doi: 10.1186/s12885-018-5036-8

Taphoorn, M. J., Claassens, L., Aaronson, N. K., Coens, C., Mauer, M., Osoba, D., et al. (2010). An international validation study of the EORTC brain cancer module (EORTC QLQ-BN20) for assessing health-related quality of life and symptoms in brain cancer patients. *Eur J Cancer, 46*(6), 1033-1040. doi: 10.1016/j.ejca.2010.01.012

Taphoorn, M. J., Dirven, L., Kanner, A. A., Lavy-Shahaf, G., Weinberg, U., Taillibert, S., et al. (2018). Influence of Treatment With Tumor-Treating Fields on Health-Related Quality of Life of Patients With Newly Diagnosed Glioblastoma: A Secondary Analysis of a Randomized Clinical Trial. *JAMA Oncology, 4*(4), 495-504. doi: 10.1001/jamaoncol.2017.5082

Taphoorn, M. J., Stupp, R., Coens, C., Osoba, D., Kortmann, R., van den Bent, M. J., et al. (2005). Health-related quality of life in patients with glioblastoma: a randomised controlled trial. *Lancet Oncol, 6*(12), 937-944. doi: 10.1016/S1470-2045(05)70432-0

Tong, M. C., Lo, P. S., Wong, K. H., Yeung, R. M., van Hasselt, C. A., Eremenco, S., et al. (2009). Development and validation of the functional assessment of cancer therapy nasopharyngeal cancer subscale. *Head Neck, 31*(6), 738-747. doi: 10.1002/hed.21023

Trinquart, L., Jacot, J., Conner, S. C., Porcher, R. (2016). Comparison of Treatment Effects

Measured by the Hazard Ratio and by the Ratio of Restricted Mean Survival Times in Oncology Randomized Controlled Trials. *J Clin Oncol, 34*(15), 1813-1819. doi: 10.1200/JCO.2015.64.2488

van Andel, G., Bottomley, A., Fossa, S. D., Efficace, F., Coens, C., Guerif, S., et al. (2008). An international field study of the EORTC QLQ-PR25: a questionnaire for assessing the health-related quality of life of patients with prostate cancer. *Eur J Cancer, 44*(16), 2418-2424. doi: 10.1016/j.ejca.2008.07.030

Velikova, G., Booth, L., Smith, A. B., Brown, P. M., Lynch, P., Brown, J. M., et al. (2004). Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. *J Clin Oncol, 22*(4), 714-724. doi: 10.1200/JCO.2004.06.078

Verma, S., O'Shaughnessy, J., Burris, H. A., Campone, M., Alba, E., Chandiwana, D., et al. (2018). Health-related quality of life of postmenopausal women with hormone receptor-positive, human epidermal growth factor receptor 2-negative advanced breast cancer treated with ribociclib + letrozole: results from MONALEESA-2. *Breast Cancer Res Treat, 170*(3), 535-545. doi: 10.1007/s10549-018-4769-z

Walters, S. J. (2009). *Quality of life outcomes in clinical trials and health-care evaluation: a practical guide to analysis and interpretation* (Vol. 84): John Wiley & Sons.

Ward, W. L., Hahn, E. A., Mo, F., Hernandez, L., Tulsky, D. S., Cella, D. (1999). Reliability and validity of the Functional Assessment of Cancer Therapy-Colorectal (FACT-C) quality of life instrument. *Qual Life Res, 8*(3), 181-195. doi: 10.1023/a:1008821826499

Ware, J. E., Jr., Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care, 30*(6), 473-483.

Ware, J. E., Jr., Snow, K. K., Kosinski, M., Gandek, B. (1993). SF-36 Health Survey. Manual and Interpretation Guide. Boston: The Health Institute, New England Medical Center.

Weis, J., Tomaszewski, K. A., Hammerlid, E., Ignacio Arraras, J., Conroy, T., Lanceley, A., et al. (2017). International Psychometric Validation of an EORTC Quality of Life Module Measuring Cancer Related Fatigue (EORTC QLQ-FA12). *J Natl Cancer Inst, 109*(5). doi: 10.1093/jnci/djw273

Weitzner, M. A., Meyers, C. A., Gelke, C. K., Byrne, K. S., Cella, D. F., Levin, V. A. (1995).

The Functional Assessment of Cancer Therapy (FACT) scale. Development of a brain subscale and revalidation of the general version (FACT-G) in patients with primary brain tumors. *Cancer, 75*(5), 1151-1161. doi: 10.1002/1097-0142(19950301)75:5<1151::aid-cncr2820750515>3.0.co;2-q

Wheelwright, S., Darlington, A. S., Fitzsimmons, D., Fayers, P., Arraras, J. I., Bonnetain, F., et al. (2013). International validation of the EORTC QLQ-ELD14 questionnaire for assessment of health-related quality of life elderly patients with cancer. *Br J Cancer, 109*(4), 852-858. doi: 10.1038/bjc.2013.407

Whistance, R. N., Conroy, T., Chie, W., Costantini, A., Sezer, O., Koller, M., et al. (2009). Clinical and psychometric validation of the EORTC QLQ-CR29 questionnaire module to assess health-related quality of life in patients with colorectal cancer. *Eur J Cancer, 45*(17), 3017-3026. doi: 10.1016/j.ejca.2009.08.014

WHO (1948). WHO Constitution. Geneva: WHO, 1948.

WHOQOL (1993). Study protocol for the World Health Organization project to develop a Quality of Life assessment instrument (WHOQOL). *Qual Life Res, 2*(2), 153-159.

Wildiers, H., Brain, E., Penninckx, B., Ring, A., Repetto, L., Soubeyran, P., et al. (2012). The EORTC Cancer in the Elderly Task Force, a Protostar for EORTC's future. *Eur J Cancer Supplements, 10*(1), 34-38.

Wilson, I. B., Cleary, P. D. (1995). Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA, 273*(1), 59-65.

Winstanley, J. B., Young, T. E., Boyle, F. M., Bergenmar, M., Bottomley, A., Burmeister, B., et al. (2015). Cross-cultural development of a quality-of-life measure for patients with melanoma: phase 3 testing of an EORTC Melanoma Module. *Melanoma Res, 25*(1), 47-58. doi: 10.1097/CMR.0000000000000122

Winters, Z. E., Afzal, M., Rutherford, C., Holzner, B., Rumpold, G., da Costa Vieira, R. A., et al. (2018). International validation of the European Organisation for Research and Treatment of Cancer QLQ-BRECON23 quality-of-life questionnaire for women undergoing breast reconstruction. *Br J Surg, 105*(3), 209-222. doi: 10.1002/bjs.10656

Winters, Z. E., Benson, J. R., Pusic, A. L. (2010). A systematic review of the clinical evidence to guide treatment recommendations in breast reconstruction based on patient-

reported outcome measures and health-related quality of life. *Ann Surg, 252*(6), 929-942. doi: 10.1097/SLA.0b013e3181e623db

Winters, Z. E., Emson, M., Griffin, C., Mills, J., Hopwood, P., Bidad, N., et al. (2015). Learning from the QUEST multicentre feasibility randomization trials in breast reconstruction after mastectomy. *Br J Surg, 102*(1), 45-56. doi: 10.1002/bjs.9690

Winters, Z. E., Griffin, C., Horne, R., Bidad, N., McCulloch, P. (2014). Barriers to accrue to clinical trials and possible solutions. *Br J Cancer, 111*(4), 637-639. doi: 10.1038/bjc.2014.318

Wood-Dauphinee, S. (1999). Assessing quality of life in clinical research: from where have we come and where are we going? *J Clin Epidemiol, 52*(4), 355-363.

Yadegarfar, G., Friend, L., Jones, L., Plum, L. M., Ardill, J., Taal, B., et al. (2013). Validation of the EORTC QLQ-GINET21 questionnaire for assessing quality of life of patients with gastrointestinal neuroendocrine tumours. *Br J Cancer, 108*(2), 301-310. doi: 10.1038/bjc.2012.560

Yost, K. J., Yount, S. E., Eton, D. T., Silberman, C., Broughton-Heyes, A., Cella, D. (2005). Validation of the Functional Assessment of Cancer Therapy-Breast Symptom Index (FBSI). *Breast Cancer Res Treat, 90*(3), 295-298. doi: 10.1007/s10549-004-5024-3

Annexes

Annexe A : Le questionnaire QLQ-C30.

EORTC QLQ-C30 (version 3)

Nous nous intéressons à vous et à votre santé. Répondez vous-même à toutes les questions en entourant le chiffre qui correspond le mieux à votre situation. Il n'y a pas de « bonne » ou de « mauvaise » réponse. Ces informations sont strictement confidentielles.

Merci de préciser :

Vos initiales : |____|____|____|

Date de naissance (jour/mois/année) : |__|__||__|__||__|__|__|__|

La date d'aujourd'hui (jour/mois/année) : |__|__||__|__||__|__|__|__|

	Pas du tout	**Un peu**	**Assez**	**Beaucoup**
1. Avez-vous des difficultés à faire certains efforts physiques pénibles comme porter un sac à provisions chargé ou une valise ?	1	2	3	4
2. Avez-vous des difficultés à faire une longue promenade ?	1	2	3	4
3. Avez-vous des difficultés à faire un petit tour dehors ?	1	2	3	4
4. Êtes-vous obligée de rester au lit ou dans un fauteuil pendant la journée ?	1	2	3	4
5. Avez-vous besoin d'aide pour manger, vous habiller, faire votre toilette ou aller aux toilettes ?	1	2	3	4

Au cours de la semaine passée :

	Pas du tout	**Un peu**	**Assez**	**Beaucoup**
6. Avez-vous été gêné(e) pour faire votre travail ou vos activités de tous les jours ?	1	2	3	4
7. Avez-vous été gêné(e) dans vos activités de loisirs ?	1	2	3	4
8. Avez-vous eu le souffle court ?	1	2	3	4
9. Avez-vous ressenti de la douleur ?	1	2	3	4
10. Avez-vous eu besoin de repos ?	1	2	3	4
11. Avez-vous eu des difficultés à dormir ?	1	2	3	4
12. Vous êtes-vous senti(e) faible ?	1	2	3	4
13. Avez-vous manqué d'appétit ?	1	2	3	4
14. Avez-vous eu des nausées (mal au cœur) ?	1	2	3	4
15. Avez-vous vomi ?	1	2	3	4

Passez à la page suivante S.V.P

Au cours de la semaine passée :	**Pas du tout**	**Un peu**	**Assez**	**Beaucoup**
16. Avez-vous été constipé(e) ?	1	2	3	4
17. Avez-vous eu de la diarrhée ?	1	2	3	4
18. Avez-vous été fatigué(e) ?	1	2	3	4
19. Des douleurs ont-elles perturbé vos activités quotidiennes ?	1	2	3	4
20. Avez-vous eu des difficultés à vous concentrer sur certaines choses, par exemple, pour lire le journal ou regarder la télévision ?	1	2	3	4
21. Vous êtes-vous senti(e) tendu(e) ?	1	2	3	4
22. Vous êtes-vous fait du souci ?	1	2	3	4
23. Vous êtes-vous senti(e) irritable ?	1	2	3	4
24. Vous êtes-vous senti(e) déprimé(e) ?	1	2	3	4
25. Avez-vous eu des difficultés à vous souvenir de certaines choses ?	1	2	3	4
26. Votre état physique ou votre traitement médical vous ont-ils gêné(e) dans votre vie familiale ?	1	2	3	4
27. Votre état physique ou votre traitement médical vous ont-ils gêné(e) dans vos activités sociales (par exemple, sortir avec des amis, aller au cinéma...) ?	1	2	3	4
28. Votre état physique ou votre traitement médical vous ont-ils causé des problèmes financiers ?	1	2	3	4

Pour les questions suivantes, veuillez répondre en entourant le chiffre entre 1 et 7 qui s'applique le mieux à votre situation :

29. Comment évalueriez-vous votre état de santé au cours de la semaine passée ?

1	2	3	4	5	6	7
Très mauvais						Excellent

30. Comment évalueriez-vous l'ensemble de votre qualité de vie au cours de la semaine passée ?

1	2	3	4	5	6	7
Très mauvais						Excellent

© Copyright 1995 EORTC Study Group on Quality of Life. Tous droits réservés. Version 3.0.

Annexe B : Le questionnaire FACT-G.

FACT-G (Version 4)

Vous trouverez ci-dessous une liste de commentaires que d'autres personnes atteintes de la même maladie que vous ont jugés importants. **Veuillez indiquer votre réponse en entourant un seul chiffre par ligne et en tenant compte des 7 derniers jours.**

	BIEN-ÊTRE PHYSIQUE	**Pas du tout**	**Un peu**	**Moyen-nement**	**Beau-coup**	**Énormé-ment**
GP1	Je manque d'énergie.....	0	1	2	3	4
GP2	J'ai des nausées	0	1	2	3	4
GP3	À cause de mon état physique, j'ai du mal à répondre aux besoins de ma famille	0	1	2	3	4
GP4	J'ai des douleurs.....	0	1	2	3	4
GP5	Je suis incommodé(e) par les effets secondaires du traitement	0	1	2	3	4
GP6	Je me sens malade	0	1	2	3	4
GP7	Je suis obligé(e) de passer du temps allongé(e)	0	1	2	3	4

	BIEN-ÊTRE SOCIAL/FAMILIAL	**Pas du tout**	**Un peu**	**Moyen-nement**	**Beau-coup**	**Énormé-ment**
GS1	Je me sens proche de mes amis	0	1	2	3	4
GS2	Ma famille me soutient moralement	0	1	2	3	4
GS3	Mes amis me soutiennent	0	1	2	3	4
GS4	Ma famille a accepté ma maladie.....	0	1	2	3	4
GS5	Je suis satisfait(e) de la communication avec ma famille au sujet de ma maladie.....	0	1	2	3	4
GS6	Je me sens proche de mon (ma) partenaire (ou de la personne qui est mon principal soutien)	0	1	2	3	4
Q1	*Quel que soit votre degré d'activité sexuelle en ce moment, veuillez répondre à la question suivante. Si vous préférez ne pas y répondre, cochez cette case et passez à la section suivante.*	☐				
GS7	Je suis satisfait(e) de ma vie sexuelle.....	0	1	2	3	4

Veuillez indiquer votre réponse en entourant un seul chiffre par ligne et en tenant compte des 7 derniers jours.

	BIEN-ÊTRE ÉMOTIONNEL	**Pas du tout**	**Un peu**	**Moyen-nement**	**Beau-coup**	**Énormé-ment**
GE1	Je me sens triste.....	0	1	2	3	4
GE2	Je suis satisfait(e) de la façon dont je fais face à ma maladie	0	1	2	3	4
GE3	Je perds espoir dans le combat contre ma maladie....	0	1	2	3	4
GE4	Je me sens nerveux (nerveuse)	0	1	2	3	4
GE5	Je suis préoccupé(e) par l'idée de mourir.....	0	1	2	3	4
GE6	Je suis préoccupé(e) à l'idée que mon état de santé puisse s'aggraver.....	0	1	2	3	4

	BIEN-ÊTRE FONCTIONNEL	**Pas du tout**	**Un peu**	**Moyen-nement**	**Beau-coup**	**Énormé-ment**
GF1	Je suis capable de travailler (y compris le travail à la maison)	0	1	2	3	4
GF2	Mon travail (y compris le travail à la maison) me donne de la satisfaction	0	1	2	3	4
GF3	Je suis capable de profiter de la vie	0	1	2	3	4
GF4	J'ai accepté ma maladie.....	0	1	2	3	4
GF5	Je dors bien.....	0	1	2	3	4
GF6	J'apprécie mes loisirs habituels.....	0	1	2	3	4
GF7	Je suis satisfait(e) de ma qualité de vie actuelle	0	1	2	3	4

Résumé

Cette thèse a pour but d'apporter une contribution à l'analyse et la comparaison des données de PROs (« *patient-reported outcomes* » ou « résultats rapportés par le patient ») dans les essais cliniques en cancérologie, dont l'interprétation reste complexe et non standardisée. Parmi les nombreuses méthodes proposées pour l'analyse longitudinale des données de PROs, figure l'approche du temps jusqu'à détérioration (TJD). Dans le cadre de ce projet, des travaux menés ont fait l'état des définitions du TJD utilisées et mettent en évidence que certaines recommandations n'ont pas été suivies. De plus, la variabilité des définitions utilisées compromet la comparaison des résultats entre les essais cliniques. L'approche du TJD nécessite une définition claire de ce qui est considéré comme une « détérioration », et cela dépendra à la fois de la localisation cancéreuse, de la situation thérapeutique, du score de référence, de la différence minimale cliniquement importante perçue par le patient, ainsi que des règles de censure. Dans l'optique d'optimiser et d'harmoniser les définitions utilisées du TJD, afin d'avoir des résultats comparables, deux macros SAS ont été créées sur la méthode du TJD afin de standardiser les définitions utilisées par la communauté. S'inscrivant dans cette trajectoire, une étude menée sur une cohorte de patientes atteintes d'un cancer du sein en situation adjuvante a conduit à s'intéresser à la première détérioration du patient et à la gestion de l'absence de randomisation à l'inclusion. En parallèle, cette méthode a également été appliquée dans un essai de phase II randomisé chez des patients atteints d'un cancer du pancréas métastatique. L'impact de l'occurrence des données manquantes à l'inclusion dans cet essai a été traité en appliquant une imputation multiple utilisant la méthode de Monte-Carlo par chaînes de Markov. Ces travaux soulignent le besoin de continuer la création de consensus pour l'analyse longitudinale des données de PROs dans les essais cliniques en cancérologie.

Mots-clés : « *patient-reported outcomes* », qualité de vie relative à la santé, analyse longitudinale, temps jusqu'à détérioration, cancérologie.

Abstract

The purpose/aim of this thesis is to contribute to the analysis and comparison of PROs (« *patient-reported outcomes* ») data in oncology clinical trials. The interpretation of such results remains complex and unstandardized. One of the many ways to carry out a longitudinal analysis of PRO data is the time to deterioration (TTD) approach. Within of the scope of this project, some of the research examined which definitions of TTD are used and pointed out that some recommendations have not been followed. Moreover, due to the variability of the definitions in use, the comparison of various results from clinical trials is compromised. A clear definition of what is considered to be a « deterioration » is required for the TTD approach. It will depend on many criteria such as the location of the cancer, the therapeutic setting, the reference score, the minimal important difference perceived by the patient, as well as on censoring rules. Two SAS macros were developed on the TTD method as a way to optimize and harmonize the TTD definitions that are being used, as well as to be able to have comparable results and consequently a way to help standardize those definitions. In this perspective, a study conducted on a cohort of adjuvant breast cancer patients led to more focus on the first deterioration of the patient and the management of non-randomization at baseline. In parallel, this method was also implemented for a randomized phase II trial on patients with metastatic pancreatic cancer. During this trial, the impact of the occurrence of missing data at baseline was handled by applying a multiple imputation based on the Markov Chain Monte Carlo method. These works highlight the need to continue developing a consensus for the longitudinal analysis of PROs data in oncology clinical trials.

Keywords: « *patient-reported outcomes* », health-related quality of life, longitudinal analysis, time to deterioration, oncology.