



**HAL**  
open science

# Les critères de jugement centrés sur le patient dans les essais cliniques en oncologie thoracique

Frédéric Fiteni

► **To cite this version:**

Frédéric Fiteni. Les critères de jugement centrés sur le patient dans les essais cliniques en oncologie thoracique. Cancer. Université de Franche-Comté, 2016. Français. NNT : 2016BESA3012 . tel-01579503

**HAL Id: tel-01579503**

**<https://theses.hal.science/tel-01579503>**

Submitted on 31 Aug 2017

**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.



**UNIVERSITE DE FRANCHE-COMTE**  
ECOLE DOCTORALE « ENVIRONNEMENTS-SANTE »

Année 2016

**THESE**

Pour obtenir le grade de

**Docteur de l'Université de Franche-Comté**

Spécialité : Méthodologie de la recherche clinique en cancérologie

Présentée et soutenue publiquement

Le 14 Octobre 2016

Par

**Frédéric Fiteni**

**Les critères de jugement centrés sur le patient dans les essais  
cliniques en oncologie thoracique**

**JURY**

<b>Pr Xavier Pivot</b> , PU-PH, Université de Franche-Comté	(Président)
<b>Pr Virginie Westeel</b> , PU-PH, Université de Franche-Comté	(directrice de thèse)
<b>Pr Franck Bonnetain</b> , PU-PH, Université de Franche-Comté	(co-directeur de thèse)
<b>Pr Alain Duhamel</b> , PU-PH, Université de Lille	(rapporteur)
<b>Pr Nicolas Girard</b> , PU-PH, Université de Lyon	(rapporteur)
<b>Pr Elisabeth Quoix</b> , PU-PH, Université de Strasbourg	(examineur)
<b>Dr Bernard Milleron</b> , Institut Francophone de Cancérologie Thoracique	(examineur)

# Remerciements

---

Je souhaite exprimer toute ma gratitude envers le Pr Franck Bonnetain. Tu m'as accueilli dans ton unité alors que je ne connaissais rien aux biostatistiques et à la méthodologie. Tu m'as fait confiance et permis de réaliser ce travail. Je t'en suis très reconnaissant et j'espère que je serais digne de ton enseignement.

Je remercie chaleureusement le Pr Virginie Westeel pour avoir accepté de diriger cette thèse. Tu m'as accueilli dans l'équipe d'oncologie thoracique avec beaucoup de gentillesse et de considération. J'ai pu apprécier ta rigueur, ta compétence et ton contact auprès des malades. Je t'en suis très reconnaissant et j'espère que je serais digne de ton enseignement.

Je remercie particulièrement le Pr Xavier Pivot pour avoir accepté de présider cette thèse ainsi que pour toutes ces années de formation en oncologie. Je vous en suis très reconnaissant et j'espère que je serais digne de votre enseignement.

Je remercie le Pr Elisabeth Quoitman sans qui ce travail aurait été impossible. Vous avez accepté de me transmettre votre base de données et vous me faites l'honneur de juger ce travail. Soyez assuré de mon profond respect et de ma sincère reconnaissance.

Je remercie l'Institut Francophone de Cancérologie Thoracique qui m'a autorisé à exploiter la base de données de l'essai IFCT 0501 et qui m'a soutenu durant tout ce travail.

Je remercie les Pr Nicolas Girard et Pr Alain Duhamel qui me font l'honneur d'être rapporteurs de cette thèse.

Je remercie le Dr Bernard Milleron qui me fait l'honneur de juger ce travail.

Je remercie Amélie Anota pour m'avoir transmis sa science de la qualité de vie. Je te souhaite beaucoup de réussite et de bonheur.

Je remercie Dewi Vernerey pour m'avoir transmis sa science des statistiques et notamment des modèles pronostiques.

Je remercie tous les membres de l'équipe de méthodologie et qualité de vie en cancérologie pour leur gentillesse et leur bonne humeur.

Je remercie tous les membres de l'équipe du service d'oncologie.

Je remercie tous mes amis ici et ailleurs.

A Alixia pour tous les bons moments passés ensemble cette année.

A ma sœur Betty. Je te souhaite beaucoup de bonheur que tu mérites tant.

A ma sœur Karla, garde toujours ta bonne humeur, ton dynamisme et ton humour. Je te souhaite beaucoup de bonheur et de réussite.

Mes pensées s'adressent tout particulièrement à mes parents qui sont des piliers sur lesquels je peux m'appuyer à tout moment. Ils m'encouragent, me soutiennent et croient en moi. C'est à eux que cette thèse est dédiée.

# Sommaire

---

Abréviations.....	6
Production scientifique.....	7
Introduction.....	9
I.	
Généralités.....	10
II. Les critères de jugement centrés sur la tumeur.....	10
1. Les critères de jugement intermédiaires centrés sur la tumeur.....	11
2. Les critères de substitution.....	14
a. Définition.....	14
b. Méthodes de validation.....	14
i. Validation sur un essai clinique.....	14
• Critères de Prentice.....	14
• La proportion expliquée de Freedman.....	15
ii. Approche méta-analytique.....	15
III. Les critères de jugement centrés sur le patient.....	16
1. La survie globale.....	17
2. La qualité de vie relative à la santé.....	17
a. La qualité de vie comme critère de jugement clinique dans les essais cliniques en cancérologie.....	19
b. Les instruments de mesure de la qualité de vie.....	20
i. Propriétés psychométriques et méthodes classiques de validation.....	20
• La validité.....	20
• La fiabilité.....	23
• La sensibilité au changement.....	24
ii. Le questionnaire QLQ-C30.....	24
iii. Le questionnaire FACT-L.....	25
c. Administration des questionnaires.....	26
d. Temps d'évaluation de la qualité de vie.....	27
e. Les dimensions cibles.....	28
f. La différence minimale cliniquement importante.....	29

g.	Analyse statistique longitudinale de la qualité de vie.....	31
i.	Population d'analyse.....	31
ii.	Méthodes d'analyse longitudinale.....	32
•	Les modèles mixtes d'analyse de variance sur mesures répétées.....	32
•	Le temps jusqu'à détérioration d'un score de qualité de vie.....	33
h.	Valeur pronostique de la qualité de vie relative à la santé sur la survie globale....	34
Objectifs.....		35
Articles.....		38
	Les critères de jugement dans les essais cliniques en cancérologie.....	39
	Les critères de jugement de type temps à évènement dans les essais cliniques de phase III des cancers bronchiques non à petites cellules avancés : une revue systématique.....	46
	Les critères de substitution de la survie globale dans les essais cliniques de cancer pulmonaire : une revue de littérature.....	60
	La qualité de vie relative à la santé dans les essais cliniques de phase III en oncologie: de l'administration du questionnaire à l'analyse statistique.....	77
	Défis dans l'analyse statistique des données de QdV dans les essais cliniques en cancérologie.....	84
	Méthodologie de l'analyse de la qualité de vie relative à la santé dans les essais cliniques de phase III des cancers bronchiques non à petites cellules avancés : une revue systématique.....	90
	La qualité de vie relative à la santé chez les patients âgés atteints d'un carcinome bronchique non à petites cellules avancé traités par le doublet de chimiothérapie carboplatine et paclitaxel hebdomadaire en comparaison à une monochimiothérapie.....	100
	La qualité de vie relative à la santé comme co-critère de jugement principal dans les essais cliniques randomisés en oncologie.....	115
	Valeur pronostique de la qualité de vie relative à la santé sur la survie globale chez les patients âgés atteints de CBNPC.....	123
Discussion.....		134

I. Les critères de jugement cliniques dans les essais de phase III de CBNPC avancés.....	135
II. Les critères de substitution de la survie globale dans les essais cliniques de cancer pulmonaire.....	135
III. Méthodologie d'analyse de la Qdv dans les essais de phase III de CBNPC avancés.....	136
IV. La qualité de vie relative à la santé chez les patients âgés atteints d'un carcinome bronchique non à petites cellules avancé traités par le doublet de chimiothérapie carboplatine et paclitaxel hebdomadaire en comparaison à une monochimiothérapie.....	137
V. La qualité de vie relative à la santé comme co-critère de jugement principal dans les essais cliniques en oncologie.....	138
VI. Valeur pronostique de la qualité de vie relative à la santé sur la survie globale chez les patients âgés atteints de CBNPC.....	138
Conclusion.....	140
Références.....	143

## *Abréviations*

---

AIC	Akaike Information Criteria
ASCO	American Society of Clinical Oncology
ADL	Activities of Daily Living
BIC	Bayesian Information Criteria
CBNPC	Carcinomes bronchiques non à petites cellules
CONSORT	Consolidated Standards of Reporting Trials
DATECAN	Definition for the Assessment of Time-to-event Endpoints in CANcer trials
DMCI	Différence minimale cliniquement importante
EMA	European Medicines Agency
EORTC	European Organization for Research and Treatment of Cancer
FACT	Functional Assessment of Cancer Therapy–General
FDA	Food and Drug Administration
HER2	Human epidermal receptor 2
IFCT	Intergroupe francophone de cancérologie thoracique
OS	Overall survival
OMS	Organisation Mondiale de la Santé
MMS	Mini Mental State
PE	Proportion expliquée
PRO	Patient reported outcome
QdV	Qualité de vie relative à la santé
RECIST	Response evaluation criteria in solid tumours
SG	Survie globale
SSP	Survie sans progression
TJD	Temps jusqu'à détérioration
TOI	Trial Outcome Index
WHO	World Health Organization
WHOQOL	World Health Organization Quality Of Life

# *Production scientifique*

---

## **Publications**

- **Fiteni F**, Westeel V, Pivot X, Borg C, Vernerey D, Bonnetain F. Endpoints in cancer clinical trials. *J Visc Surg.* févr 2014;151(1):17-22.
- **Fiteni F**, Anota A, Westeel V, Bonnetain F. [Health-related quality of life in phase III cancer clinical trials: From questionnaire administration to statistical analysis]. *Bull Cancer (Paris).* avr 2015;102(4):360-6.
- **Fiteni F**, Pam A, Anota A, Vernerey D, Paget-Bailly S, Westeel V, et al. Health-related quality-of-life as co-primary endpoint in randomized clinical trials in oncology. *Expert Rev Anticancer Ther.* 31 mai 2015;1-7.
- **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(1):122.
- **Fiteni F**, Vernerey D, Bonnetain F, Vaylet F, Sennéart H, Trédaniel J, et al. Prognostic value of health-related quality of life for overall survival in elderly non-small-cell lung cancer patients. *Eur J Cancer Oxf Engl 1990.* janv 2016;52:120-8.
- 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 Off J Am Soc Clin Oncol.* 18 avr 2016;
- **Fiteni F**, Anota A, Bonnetain F, Oster J-P, Pichon E, Wislez M, et al. Health-related quality of life in elderly patients with advanced non-small cell lung cancer comparing carboplatin and weekly paclitaxel doublet chemotherapy with monotherapy. *Eur Respir J.* 23 juin 2016;



## Articles soumis

- **Fiteni F**, Mouillet G, Vernerey D, Westeel, Bonnetain F. Time-to-event Endpoints Reporting in Phase III Advanced Non-Small-Cell Lung Cancer Clinical Trials: A Systematic Review
- **Fiteni F**, Westeel V, Bonnetain F. Surrogate endpoints for overall survival in lung cancer trials: a review

## *Introduction*

---

« Savoir pour prévoir, afin de pouvoir. »

*Auguste Comte*

## **I. Généralités**

Les critères de jugement sont des mesures cliniques et biologiques pour l'évaluation et le développement des thérapeutiques. Les critères de jugement font référence à des mesures cliniques ou biologiques afin d'évaluer l'efficacité de stratégies thérapeutiques. L'American Society of Clinical Oncology (ASCO) considère que le but d'un traitement actif en oncologie est d'améliorer la quantité et/ou la qualité de vie des patients. Les critères de jugement cliniques primaires et secondaires peuvent être classés en deux catégories principales : les « critères de jugement centrés sur le patient » comprenant la survie globale (SG) et la qualité de vie relative à la santé (QdV) et les « critères de jugement centrés sur la tumeur » comme la survie sans progression. Les critères de jugement peuvent être des critères composites c'est-à-dire constitués de plusieurs événements. Plusieurs critères de jugement peuvent être associés comme co-critères principaux. La SG en tant que critère de jugement principal est la référence pour démontrer un bénéfice clinique au patient mais avec l'augmentation du nombre de traitements efficaces disponibles dans beaucoup de cancers (i.e. les différentes lignes de traitement), il est nécessaire d'inclure un nombre plus important de patients et de les suivre plus longtemps afin d'observer le nombre d'événements nécessaires pour atteindre la puissance statistique. Ceci augmente le coût des essais cliniques et le temps avant l'obtention des résultats. Par conséquent, les critères de jugement centrés sur la tumeur comme la survie sans progression sont souvent utilisés car ils sont évalués plus précocement (i.e. critères intermédiaires). Cependant, ils n'ont pas de définitions consistantes et ne sont pas systématiquement validés comme critères de substitution de la survie globale. Dans ce contexte, la QdV constitue un critère de jugement alternatif qui évalue un bénéfice clinique direct au patient.

## **II. Les critères de jugement centrés sur la tumeur**

Les critères de jugement centrés sur la tumeur sont des marqueurs biologiques. La plupart du temps ces critères de jugement sont des réponses pharmacologiques à une intervention thérapeutique. Par exemple, la réponse tumorale selon RECIST, le taux de cellules tumorales circulantes, la survie sans maladie et la survie sans progression sont des critères de jugement centrés sur la tumeur. Néanmoins, ils ne reflètent pas un bénéfice direct pour le patient. Les conclusions des essais cliniques qui utilisent ce genre de critères sont remises en question

quant au bénéfice réel, les agences de régulation demandent souvent une analyse complémentaire de la QdV si celle-ci n'a pas été réalisée. Leur validation comme critère de substitution à des critères centrés sur le patient est un préalable nécessaire.

### **1. Les critères de jugement intermédiaires centrés sur la tumeur**

Les critères de jugement centrés sur la tumeur sont généralement des critères composites, ils sont basés sur l'évaluation de la tumeur et combinent différents événements comme la progression locale et à distance, la récurrence locale et à distance, la survenue d'un second cancer, le décès ou une toxicité sévère. En général, la survie sans progression est le temps entre la randomisation et la progression tumorale ou le décès quelle que soit la cause à la phase métastatique. La survie sans maladie est en général définie comme le temps entre la randomisation et la récurrence tumorale ou le décès quelle que soit la cause après un traitement à visée curatif. Ces critères de jugement sont fréquemment utilisés par les oncologues car ils sont disponibles plus précocement que la SG, moins influencés par les causes compétitives de décès et ne dépendent pas des traitements administrés après la progression. Néanmoins, ils présentent d'importantes limites :

- il existe une variété de critères composites utilisés comme critères de jugement principaux dans les essais cliniques en oncologie. Le temps jusqu'à progression est différent de la survie sans progression, l'évènement d'intérêt est seulement la progression de la maladie, les patients qui décèdent quelle que soit la cause sont censurés. La survie sans progression et le temps jusqu'à progression sont utilisés de manière interchangeable dans la littérature récente. Par conséquent, la comparaison des résultats des différents essais cliniques est difficile.
- il n'existe aucun consensus international sur la définition de la survie sans progression et la survie sans maladie, par conséquent il existe une variabilité de la définition d'un même critère de jugement dans différents essais cliniques qui étudient la même maladie. Par exemple, les 4 essais randomisés étudiant le trastuzumab dans les cancers du sein HER2 positifs en situation adjuvante illustrent cette confusion qui peut émerger d'une définition hétérogène de la survie sans maladie (Tableau 1) (1–4). De plus, Birgisson et al. (5) ont démontré que l'inclusion d'un second cancer primitif différent du cancer

colorectal initial comme évènement de la définition de la survie sans maladie influence les résultats. Le taux de survie sans maladie de patients atteints d'un cancer colique stade I à III était de 68 % si cet évènement était pris en compte, comparé à 58 % s'il ne l'était pas. La différence était plus importante pour les stades II (68 % versus 60 %) que pour les stades III (49 % versus 47 %). Un autre exemple provient de l'essai PETACC 03 (6) qui était significatif ou qui ne l'était pas selon qu'un second cancer primitif était compté comme évènement dans la définition de la survie sans maladie (7);

- la détermination de la mesure tumorale de référence peut être celle avant le début du traitement ou la meilleure réponse tumorale. Dans les critères RECIST 1.1, la progression est définie comme une augmentation d'au moins 20 % de la somme des diamètres des lésions cibles en prenant comme référence la plus petite somme de l'étude (ce qui inclut la mesure à l'inclusion si elle est la plus petite de l'étude) ;
- il existe une variabilité de la définition de la progression (clinique, radiologique [RECIST] ou biologique). Cela complique la comparaison des résultats entre les études. De plus la progression clinique non radiologique est souvent subjective, avec un manque de définition standardisée pouvant constituer une source potentielle de biais (8);
- la progression est évaluée par des examens radiologiques à des intervalles réguliers mais la vraie progression se trouve en réalité entre deux évaluations. De plus ; les intervalles de surveillance sont hétérogènes inter- et intra-essais (9);
- l'évaluation radiologique de la progression est sujette à des erreurs de mesure et des biais ;
- Peu de critères de intermédiaires n'ont été validés comme critères de substitution de la SG (10,11). Néanmoins, la survie sans progression est souvent le critère de jugement principal dans d'autres types de cancers et de nombreux essais démontrent une amélioration de la survie sans progression sans amélioration de la SG. Par exemple, l'essai E2100 trial (12) a étudié l'association bevacizumab au paclitaxel hebdomadaire dans les cancers du sein métastatiques et a montré une amélioration de 6 mois de la médiane de la survie sans progression par rapport au paclitaxel seul. Par conséquent, la FDA

a accéléré son autorisation de mise sur le marché. Des études ultérieures en double insu avec d'autres chimiothérapies associées au bévacizumab ont confirmé une amélioration statistiquement significative de la survie sans progression mais plus modeste. Aucune d'entre elles n'a démontré d'amélioration de la SG. Au vu de l'absence de bénéfice en survie globale et compte tenu de la toxicité, la FDA a pris la décision inhabituelle de retirer l'indication du bevacizumab dans les cancers du sein métastatiques. Inversement, the European Medicines Agency (EMA) a approuvé sa prescription en association avec le paclitaxel et avec la capécitabine (13). Cela démontre que l'intérêt clinique de ce critère de jugement dans les essais cliniques en cancérologie est de plus en plus remis en question et demeurera questionnable.

La variété des critères de jugement composites de type temps jusqu'à évènement et la variabilité de leurs définitions, notamment de la survie sans progression et de la survie sans maladie, sont reconnues comme un problème méthodologique majeur et les recherches doivent être poursuivies pour améliorer leur fiabilité et leur reproductibilité.

Trials	Events used to define DFS as primary time-to-event endpoint							
	Local/regional recurrence	Distant metastases	Second primary cancer	Controlateral invasive cancer	Death from any causes	Death before recurrence or a second primary cancer	Ipsilateral DCSI	Controlateral DCSI
HERA	X	X	X	X		X		
NSABP B-31/NCCTG N9831	X	X		X		X		X
BCIRG 006	X	X	X		X			
FinHer		X		X				

**Tableau 1. Exemple de variabilité de définition de la survie sans maladie dans les 4 essais randomisés évaluant le trastuzumab dans les cancers du sein HER2 positifs en situation adjuvante**

## **2. Les critères de substitution**

### **a. Définition**

Un critère de substitution est un critère intermédiaire (évalué plus précocement que la SG) centré sur la tumeur dont l'objectif est de se substituer à un critère de jugement centré sur le patient.

En oncologie, les critères intermédiaires qui pourraient être utilisés comme critères de substitution à la survie globale sont de plus en plus étudiés. Néanmoins leur validation statistique en tant que critère de substitution est nécessaire. La validation des critères de substitution doit se faire pour une classe thérapeutique spécifique, dans une maladie spécifique à un stade spécifique.

### **b. Méthodes de validation**

La validation statistique de critères de substitution a fait l'objet de nombreuses recherches méthodologiques ces dernières années, mais aucun consensus n'existe sur les méthodes à utiliser pour identifier des critères de substitution. Cependant, deux grands types d'approches se dégagent : celles basées sur un essai clinique et la méthode méta-analytique.

#### **i. Validation sur un essai clinique**

##### **• Critères de Prentice**

En 1989, Prentice expliqua qu'un critère de substitution devrait « apporter une information claire de l'effet différentiel du traitement sur le vrai critère de jugement » et définit un critère de substitution valide comme « une variable de réponse pour laquelle un test d'hypothèse nulle d'absence de relation à un groupe de traitement est aussi un test valide d'hypothèse nulle correspondante basée sur le vrai critère de jugement » (14). Prentice a proposé 4 conditions de validation :

- l'effet du traitement statistiquement significatif sur le critère de substitution ;
- l'effet du traitement statistiquement significatif sur la SG ;
- l'effet pronostique statistiquement significatif du critère de substitution ;

- après ajustement sur le critère de substitution, l'effet du traitement sur la SG n'est plus statistiquement significatif : le critère de substitution capture complètement l'effet du traitement.

Cependant, la principale limite à cette méthode est la démonstration de la dernière condition, le rejet de l'hypothèse nulle peut être simplement due à un manque de puissance.

- **La proportion expliquée de Freedman**

Freedman et al. ont proposé de quantifier la part de l'effet traitement sur le critère principal expliquée par le critère de substitution (15). La proportion expliquée (PE) se calcule par :

$$PE = 1 - (\beta_S / \beta)$$

où  $\beta$  et  $\beta_S$  sont les estimations de l'effet traitement sur le critère principal, sans et avec ajustement sur le critère de substitution, respectivement.

La limite principale de cette technique est que PE peut être négative ou supérieure à 1, rendant son interprétation délicate pour une proportion.

## ii. Approche méta-analytique

La méthode méta-analytique est désormais la référence pour valider un critère de substitution. Basée sur les résultats d'essais publiés et en utilisant des données individuelles, cette approche estime conjointement :

- la corrélation entre le critère de substitution et le critère principal au niveau individuel
- la corrélation entre les effets du traitement sur chacun des deux critères.

Si le critère de substitution doit en premier lieu être un marqueur pronostique, sans référence à une intervention thérapeutique spécifique il doit être également prédictif de l'effet du traitement. Le premier aspect de la substitution est appelé « individual-level surrogacy », ce qui signifie que pour un patient particulier, le critère de substitution doit être corrélé au critère de jugement final d'intérêt comme la SG. Dans un second temps, l'effet du traitement sur le critère de substitution candidat doit être corrélé à l'effet du traitement sur le vrai critère de jugement. Cet aspect de la substitution est appelé « trial-level surrogacy » puisque cela doit être démontré pour un groupe de patients avec comme unité d'analyse statistique l'essai clinique.



### **III. Les critères de jugement centrés sur le patient**

Les critères de jugement centrés sur le patient sont des variables qui mesurent comment un patient perçoit son état de santé ou survit (16). Ils évaluent l'effet d'une intervention thérapeutique sur le patient dans les essais cliniques. De tels critères de jugement comprennent dans les essais cliniques en cancérologie la SG et la QdV. Ils reflètent ainsi un bénéfice clinique direct pour le patient (17,18).

#### **1. La survie globale**

La SG est considérée comme le critère de jugement principal de référence et le plus pertinent dans les essais cliniques en oncologie. La US Food and Drug Administration (FDA) estime la SG comme étant une mesure unanimement admise pour évaluer le bénéfice direct du traitement et doit être le critère de jugement privilégié pour les essais de phase III. Dans son rapport concernant spécifiquement les carcinomes bronchiques non à petites cellules (CBNPC) (Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics Guidance for Industry) la FDA précise: "Overall survival (OS) is an optimal endpoint because the measurement is accurate, is observed on a daily basis, and provides direct evidence of clinical benefit to the patient. Regular approval was granted on the basis of a significant improvement in OS" (19).

La SG est définie comme le temps entre la randomisation (ou l'inclusion pour les études de cohorte) ou le début du traitement jusqu'au décès quelle que soit la cause. Les patients perdus de vue ou vivants à la date de point sont censurés. Un autre critère de jugement est la survie spécifique au cancer qui est définie comme le temps entre la randomisation ou le début du traitement jusqu'au décès liés au cancer qu'il soit lié à la tumeur primitive ou à un second cancer. Les décès dus à un autre cancer, les décès non liés au cancer, les décès liés au traitement, les perdus de vue et les patients vivants à la date de point sont censurés. La difficulté est d'identifier la cause spécifique du décès en l'absence d'autopsie. Par ailleurs, ce critère de jugement est de nature composite d'où la difficulté de le définir.

À l'heure actuelle, avec l'augmentation du nombre de traitements efficaces disponibles pour de nombreuses localisations cancéreuses, un nombre plus important de patients inclus et un suivi plus long sont nécessaires afin d'observer le nombre de décès permettant d'atteindre la puissance statistique requise pour démontrer une amélioration de la SG. Dans ce contexte, nous pouvons être confrontés à un risque plus important de perdus de vue (biais

d'information). De plus, pour les cancers ayant un bon pronostic, la différence attendue de SG entre les 2 bras de traitement est petite, il est donc nécessaire d'inclure de nombreux patients afin d'observer le nombre de décès attendu. De ce fait, l'augmentation du nombre de patients inclus et de la durée de leur suivi augmente le coût des essais cliniques. Par ailleurs, des thérapies efficaces sont disponibles après la participation à l'essai clinique, par conséquent, l'effet des traitements étudiés sur la SG peut se confondre avec l'effet des traitements administrés après l'essai. Les cancers du sein illustrent ce problème; à ce jour, il devient de plus en plus difficile de démontrer un bénéfice en SG en 1re ligne métastatique probablement en raison de l'administration après l'essai de thérapies efficaces si ces dernières ne sont pas contrôlées (20). Néanmoins, étant donné la randomisation, nous pouvons supposer que des stratégies de prise en charge standards sont poursuivies après l'essai et distribuées de manière homogène entre les bras de traitement supprimant cet effet de confusion. Pour l'ensemble de ces raisons, de plus en plus de critères de jugement qui sont évalués plus précocement que la SG sont utilisés dans les essais cliniques en cancérologie.

## **2. La qualité de vie relative à la santé**

En 1948, l'Organisation Mondiale de la Santé (OMS) a défini la santé comme « un état de complet bien-être physique, mental et social, et pas seulement l'absence de maladie ou d'infirmité » (WHO, 1948).

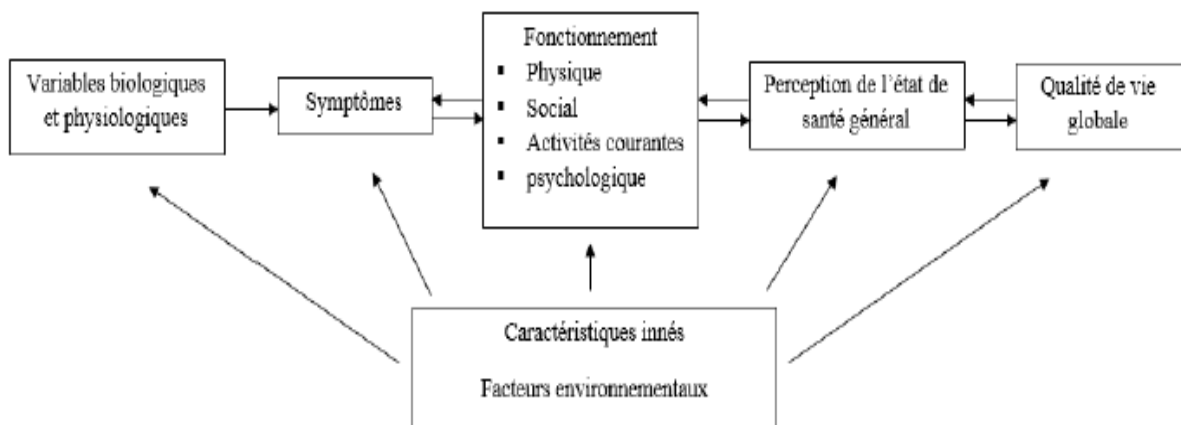
En 1993, l'OMS définit la qualité de vie comme « la perception qu'a un individu de sa place dans l'existence, dans le contexte de la culture et du système de valeurs dans lequel il vit, en relation avec ses objectifs, ses attentes, ses normes et ses inquiétudes. Il s'agit donc d'un large champ conceptuel,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

Bien qu'il n'existe pas de consensus autour de la définition de la QdV, elle est généralement considérée comme un concept multidimensionnel qui inclut au minimum le bien-être physique, psychologique et social mais aussi les symptômes liés à la maladie et aux traitements.

La QdV entre dans le champ des « Patient-Reported Outcomes » (PROs), i.e. des mesures de l'état de santé perçue par le patient (21). Ces mesures doivent donc être rapportées par le

patient lui-même. Les PROs peuvent correspondre à une large variété de paramètres comme les symptômes liés à la maladie ou au traitement, par exemple, la fatigue ou la douleur. La satisfaction vis-à-vis des soins entre également dans le champ des PROs.

L'évaluation de la QdV est influencée par un certain nombre de paramètres. Bien qu'il n'existe pas de consensus autour de la définition de la QdV, un modèle conceptuel proposé par Wilson et Cleary en 1995 a souvent servi d'illustration pour décrire les différents facteurs influençant potentiellement l'évaluation de la QdV et les différentes interactions possibles avec des paramètres extérieurs (22). Le modèle ainsi décrit est un modèle en cinq niveaux. Le premier niveau correspond aux critères biologiques et physiologiques. Ceux-ci ont un impact sur le niveau symptomatique du patient, caractérisant la perception que le patient a de ses symptômes. Le troisième niveau correspond à la perception que le patient a de son état fonctionnel, incluant son état physique et émotionnel. Ce troisième niveau est en lien direct avec la perception que le patient a de son état de santé globale correspondant au quatrième niveau. Cette perception de sa santé globale va enfin jouer un rôle direct sur son évaluation de sa QdV globale. Dans ce modèle, le lien entre chaque niveau est unidirectionnel. En 2007, Osoba et al. (23) ont proposé une version révisée de ce schéma dans laquelle des interactions dans les deux sens sont possibles entre l'état symptomatique, l'état fonctionnel, la santé perçue et la qualité de vie globale (Figure 1).



**Figure 1.** Modification du modèle de Wilson et Cleary indiquant le potentiel d'une interaction à double sens entre plusieurs composantes d'un modèle

**a. La qualité de vie comme critère de jugement clinique dans les essais cliniques en cancérologie**

La QdV est reconnue par l'ASCO et la FDA comme un critère de jugement des essais cliniques en oncologie (24). Comme nous l'avons déjà dit le critère de jugement de référence pour évaluer un nouveau traitement dans un essai de phase III est la SG. Cependant, avec l'augmentation des thérapies efficaces contre beaucoup de cancers, un suivi plus long et un nombre plus important d'inclusions sont nécessaires pour observer un nombre de décès suffisants et atteindre la puissance statistique requise (20). Par conséquent, de nombreux essais cliniques utilisent des critères de jugement intermédiaires évaluables plus précocement et qui requièrent moins de patients comme la survie sans progression (SSP). Néanmoins, ceux-ci n'ont pas été validés comme critères de substitution de la SG et présentent une hétérogénéité de définition entre les essais cliniques (25). Dans ce contexte, les résultats de QdV dans les essais de phase III peuvent aider les instances à la prise de décision d'autorisation de mise sur le marché de nouveau traitement et aider les cliniciens à la prise de décision thérapeutique. Cependant, du fait de son caractère subjectif et multidimensionnel, son évaluation, analyse et interprétation des résultats demeurent complexes.

La FDA souligne l'importance de la mesure de la QdV dans les essais cliniques des CBNPC : "Patient-reported outcome (PRO) measures of tumor-related symptoms and functioning can provide direct evidence of treatment benefit if demonstrated to be well-defined and reliable assessments of a clinically meaningful concept or set of concepts, and if evaluated in well-conducted, placebo-controlled or double-blinded, randomized trials" (19).

La QdV est la plupart du temps un critère de jugement secondaire dans les essais de phase III. De plus en plus d'auteurs publient dans un second temps l'analyse de QdV, par exemple, pour les cancers du sein : BOLERO-2 (26,27), EMILIA (28,29), CLEOPATRA (30,31); les cancers broncho-pulmonaires: LUX- lung 3 (32,33), OPTIMAL (34,35), AVAPERL (36,37); les cancers ovariens : AURELIA (38,39), GOG 0218 (40,41); les cancers pancréatiques : PRODIGE 4/ACCORD 11 (42,43).

Un design alternatif serait d'associer la QdV et un critère de jugement intermédiaire (par exemple la SSP) comme co-critères de jugement principaux, notamment lorsque le traitement est responsable de toxicités et que le critère intermédiaire n'est pas un critère de substitution de la survie globale. Par exemple Hussain et al. (44) ont comparé une privation androgénique intermittente et continue chez des patients atteints de cancer de prostate métastatique. Les co-critères de jugement étaient d'évaluer si l'hormonothérapie intermittente est non-inférieure à

l'hormonothérapie continue en ce qui concerne la survie globale et si la QdV est différente entre les 2 groupes. C Von Plessen et al. (45) ont comparé 3 versus 6 cycles de chimiothérapie à base de sels de platines sur la survie globale et la QdV chez des patients atteints de CBNPC de stade IIIB et IV. Cette approche est récente et des recherches méthodologiques doivent être poursuivies pour définir clairement les règles de décision de tels designs.

## **b. Les instruments de mesure de la qualité de vie**

Généralement, la QdV est mesurée par le biais de questionnaires administrés aux patients. De nombreux questionnaires de QdV et de PROs ont ainsi été développés. Ces questionnaires doivent respecter un certain nombre de propriétés psychométriques pour pouvoir être validés et utilisés avec fiabilité. Ils doivent être adaptés à la population d'étude et doivent être validés dans la langue d'utilisation.

Ces questionnaires peuvent être génériques ou bien spécifiques d'une maladie. Les questionnaires génériques peuvent être administrés à tout individu quel que soit son état de santé. Les questionnaires spécifiques d'une maladie ont en revanche été développés pour une pathologie donnée et mesurent ainsi des domaines de QdV potentiellement impactés par cette pathologie.

Dans les essais de phase III en oncologie, les deux principaux questionnaires utilisés sont l'European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire QLQ-C30 et le Functional Assessment of Cancer Therapy–General (FACT-G) (46,47).

### **i. Propriétés psychométriques et méthodes classiques de validation**

- **La validité**

La validité d'un questionnaire correspond à la capacité de ce questionnaire à mesurer le (ou les) domaine(s) de QdV qu'il est censé mesurer.

**La validité apparente** (« face validity ») est une première mesure, davantage qualitative, de la validité du questionnaire. Elle permet de s'assurer que les items couvrent bien a priori les domaines qu'ils sont censés mesurer et sans ambiguïté. La « face validity » est généralement mesurée par le biais d'une évaluation par les patients et les experts médicaux de la validité du

questionnaire. Elle est souvent évaluée auprès d'un sous-échantillon de patients au moment du pré-test du questionnaire. On recueille alors l'avis du patient vis-à-vis du questionnaire, soit lors d'un entretien avec un attaché de recherche clinique par exemple, soit par le biais d'un questionnaire de débriefing auto-administré. Cette évaluation porte principalement sur la compréhension et la formulation des questions, leur redondance et leur pertinence. On demande par exemple aux patients si le questionnaire leur a semblé trop long ou trop compliqué, si certaines questions sont dérangeantes, confuses, redondantes ou au contraire si certains aspects importants de leur QdV, certains effets du traitement ou symptômes de la maladie n'ont pas été abordés.

La « face validity » correspond également à une mesure de l'acceptabilité du questionnaire par le patient : pour qu'un questionnaire puisse être utilisé, il doit être accepté par le patient, i.e. que ce dernier doit accepter de remplir le questionnaire proposé. Les questions doivent donc être compréhensibles et ne doivent être ni confuses ni dérangeantes pour le patient. Un questionnaire trop long, trop compliqué ou redondant pourrait ainsi lasser le patient, ce qui le conduirait à ne pas remplir l'intégralité du questionnaire.

L'acceptabilité d'un questionnaire est ainsi mesurée par :

- Le taux de participation des patients ;
- le taux de complétion des questionnaires retournés : les questionnaires doivent présenter peu de questions ou de dimensions manquantes ;
- et le temps de remplissage du questionnaire.

Des taux élevés de participation et de complétion des questionnaires permettront de s'assurer de l'absence d'un biais de sélection de populations. En effet, un mauvais taux de complétion pourrait conduire à mener l'analyse sur une sous-population de patients en meilleur état de santé et/ou une population particulière telle que des patients plus jeunes que la population initialement ciblée.

L'acceptabilité d'un questionnaire est également illustrée par l'absence d'effet plafond et/ou d'effet plancher. L'effet plafond est illustré par un nombre élevé de patients ayant obtenu le plus haut score qu'un patient puisse obtenir pour une dimension donnée en déclarant par exemple un niveau de symptôme maximal ou un niveau fonctionnel maximal. Inversement, l'effet plancher correspond à un taux élevé de patients ayant obtenu le score le plus bas qu'un patient puisse obtenir pour une dimension donnée. L'existence d'effet plafond et/ou d'effet plancher pour une dimension donnée sous-entendrait que les modalités de réponse aux items ne sont pas suffisantes ou bien que les items eux-mêmes ne sont pas adaptés à la population

d'étude. Le pourcentage de patients ayant obtenu les scores les plus faibles et les plus élevés possibles pour une dimension donnée doivent donc être limités (sans pour autant être nul).

**La validité de contenu** (« content validity ») concerne l'adéquation du contenu d'un instrument en termes de nombre d'items et de leur champ/portée. La validité de contenu est assez proche de la validité apparente et ces deux aspects peuvent être confondus. La différence entre les deux types de validité est que la validité apparente correspond à une revue critique des items après le développement du questionnaire, alors que la validité de contenu est réalisée au cours du développement du questionnaire et consiste à s'assurer que celui-ci couvre toutes les questions importantes.

**La validité de structure** (« construct validity ») est une des caractéristiques les plus importantes de l'instrument de mesure. Elle correspond à une évaluation de la mesure dans laquelle le questionnaire mesure bien le concept qu'il est censé mesurer. Elle permet de vérifier :

- La dimensionnalité du questionnaire : est-ce que tous les items d'une même sous-échelle sont liés à une seule variable latente ou bien davantage de variables latentes sont-elles nécessaires pour expliquer la variabilité observée ?
- L'homogénéité des items: est-ce que tous les items d'une même sous-échelle sont bien fortement liés à cette échelle ?
- Le chevauchement éventuel entre les différents domaines : est-ce que certains items d'une sous-échelle donnée sont également liés à une autre échelle du questionnaire ?

La validité de structure se décompose généralement en trois types de validité : la validité convergente, la validité divergente et la validité de groupes connus.

**Les validités convergente et divergente** sont généralement évaluées par le biais de la matrice de corrélation « multi-traités multi-méthodes ». Ainsi, les items d'une même sous-échelle sont censés être davantage corrélés aux autres items de la même sous-échelle qu'aux autres domaines du questionnaire (matrice multi-traités). Si les items sont également comparés aux domaines d'un autre questionnaire, on parle de matrice multi-méthodes. La validité convergente est évaluée à l'aide du coefficient de corrélation de Spearman entre l'item et le score obtenu à sa propre échelle, calculé en excluant l'item considéré. Concernant la validité divergente, la corrélation entre chaque item et le score obtenu à sa propre échelle est censée

être plus élevée que la corrélation entre l'item et les scores obtenus pour les autres échelles du questionnaire.

Les analyses factorielles de type Analyses Factorielles des Correspondances Multiples (AFCM) ou Analyses en Composantes Principales (ACP) permettent également de valider la structure interne du questionnaire, i.e. de vérifier que les items d'une même dimension sont bien davantage liés entre eux qu'aux items des autres dimensions.

**La validité de groupes connus** (« known-group validity ») correspond à la capacité du questionnaire à mettre en évidence une différence entre deux groupes de patients ne présentant pas le même état de santé. Par exemple, on pourrait s'attendre à ce que des patients avec un cancer à un stade avancé aient un niveau de QdV plus bas que ceux ayant un cancer à un stade précoce. Un instrument de QdV valide est censé mettre en évidence une différence de niveau de QdV entre ces deux groupes.

**La validité de critère** (« criterion validity ») consiste à comparer l'instrument en cours de développement à un instrument de référence (déjà validé) qui est censé mesurer le même concept ou les mêmes dimensions de QdV. Une matrice de corrélation entre les scores des différents questionnaires permet ainsi d'apprécier l'adéquation des dimensions censées mesurer le même domaine de QdV.

- **La fiabilité**

La fiabilité correspond à la capacité d'un instrument de mesure à donner des résultats reproductibles et cohérents.

La cohérence interne des questionnaires est généralement appréciée à l'aide du coefficient alpha de Cronbach. Ce coefficient varie entre 0 et 1. Les items d'une dimension donnée présentent une bonne cohérence interne si le coefficient alpha de Cronbach est au moins égal à 0.70.

La reproductibilité (ou répétabilité) du questionnaire peut être mesurée selon la méthode « test-retest ». Cette méthode nécessite que les patients complètent deux fois le questionnaire. L'intervalle de temps entre les deux mesures doit être relativement court (environ une à deux semaines maximum) et le patient doit être dans un état de santé stable entre les deux passations. L'intervalle de temps entre les deux passations ne doit pas non plus être trop court



puisque le patient pourrait se souvenir des réponses données lors de la première passation. La reproductibilité est alors généralement mesurée par le coefficient de corrélation intra-classe (CCI). Le CCI mesure la force de l'accord entre des mesures répétées. Il s'applique à des données continues, mais peut également être appliqué à des données catégorielles ordinales avec au moins 4 modalités de réponse. Il mesure la proportion de la variance totale qui est associée avec la variabilité inter-patient. Un coefficient supérieur à 0.90 correspondra à une très bonne reproductibilité de l'échelle et supérieur à 0.70 à une bonne reproductibilité.

- **La sensibilité au changement**

La sensibilité au changement est la capacité d'un questionnaire à déceler un changement de l'état de santé du patient sans qu'il soit nécessairement cliniquement pertinent pour le patient. Différentes méthodes statistiques existent pour déterminer la sensibilité au changement d'un questionnaire. Il n'existe pas de consensus sur les méthodes statistiques à utiliser. Néanmoins, les deux indicateurs les plus couramment utilisés sont l'effet taille (« Effect Size ») (48) et la réponse moyenne standardisée (« Standardized Response Mean ») (49).

L'« Effect Size » correspond au changement moyen entre les deux temps de mesure et divisé par l'écart type du premier temps de mesure. Le coefficient « Standardized Response Mean » correspond au changement moyen entre les deux temps de mesure divisé par l'écart type de la différence

Pour ces indicateurs, plus le coefficient est élevé, plus le changement est important: un coefficient inférieur 0.2 en valeur absolue est considéré comme un changement non significatif, entre 0.2 et 0.5 comme modéré et supérieur à 0.5 comme large.

## **ii. Le questionnaire QLQ-C30**

Le groupe QdV de l'EORTC (<http://groups.eortc.be/qol/quality-life-group>) a développé et validé un questionnaire de QdV appelé QLQ-C30 spécifique du cancer (46). Ce questionnaire contient 30 items et permet d'évaluer 15 dimensions de QdV (Annexe A) :

- cinq échelles fonctionnelles : physique, rôle, émotionnelle, cognitive et sociale,
- une dimension de QdV/santé globale,
- huit échelles symptomatiques : fatigue, nausée et vomissement, douleur, dyspnée, insomnie, perte d'appétit, constipation et diarrhée,

- ainsi que les difficultés financières liées à la maladie.

Les 28 premiers items sont construits sur une échelle de Likert à 4 modalités de réponse de type : 1 « Pas du tout » / 2 « Un peu » / 3 « Assez » / 4 « Beaucoup ». Les deux derniers items sont construits sur une échelle à 7 modalités de réponse. Ces deux items évaluent respectivement l'état physique et la QdV globale du patient ; la modalité de réponse 1 correspondant à un état « très mauvais » et la modalité de réponse 7 à un « excellent » état.

Un manuel de scoring a été créé par l'EORTC afin de donner les principales recommandations de calcul des scores. Un score est ainsi calculé par dimension et standardisé de 0 à 100 de sorte qu'un score faible corresponde à un faible niveau fonctionnel, un faible niveau de QdV/santé globale et à une absence de symptôme ; et un score élevé corresponde à un haut niveau fonctionnel, un haut niveau de QdV/santé globale et à une présence élevée de symptômes. Pour qu'un score puisse être estimé pour une dimension donnée, l'EORTC recommande qu'au moins 50% des items de la dimension considérée soient renseignés. Par défaut, la méthode d'imputation des items manquants proposée par l'EORTC est la méthode d'estimation simple par la moyenne, en considérant donc que l'item manquant ne diffère pas significativement des items renseignés. L'item manquant est alors remplacé par la moyenne des items répondus pour un individu donné.

Selon les localisations cancéreuses et les modalités de prise en charge, différents modules supplémentaires ont été développés par l'EORTC. Ces modules sont à administrer conjointement avec le questionnaire QLQ-C30. Ainsi le QLQ-LC13 pour le cancer du poumon contient 13 items et 10 dimensions (dyspnée, toux, hémoptysie, douleurs buccales, neuropathie périphérique, alopecie, douleur dans la poitrine, douleur dans le bras ou l'épaule, autre douleur).

#### **i. Le questionnaire FACT-G**

Le groupe américain « Functional Assessment of Cancer Therapy » (FACT) (<http://www.facit.org/>) a également développé et validé un questionnaire de QdV appelé FACT-General (FACT-G) spécifique du cancer (47). Ce questionnaire contient 27 items construits sur une échelle de Likert à 5 modalités de réponses codées : 0 « Pas du tout » / 1 « Un peu » / 2 « Moyennement » / 3 « Beaucoup » / 4 « Enormément ». Quatre dimensions de bien-être sont évaluées par le biais de ces items : le bien-être physique, social/familial, émotionnel et fonctionnel.

Comme pour les questionnaires de l'EORTC, un score est calculé par dimension. Ces scores varient de 0 à 28 où 0 représente un bas niveau de bien-être et 28 un haut niveau de bien-être. Pour qu'un score puisse être estimé, au moins 50% des items de la dimension doivent être renseignés. Contrairement au questionnaire QLQ-C30, les items du questionnaire FACT-G sont séparés par dimension et le nom de chaque dimension évaluée précède les items correspondant. De plus, un score FACT-G global est estimé à partir des scores obtenus pour chaque dimension. Ce score correspond à la somme des scores bruts obtenus pour chaque dimension et varie de 0 (faible niveau) à 108 (bon niveau de bien-être général). Ce score global ne peut être calculé que si l'ensemble des sous-scores ont pu être estimés. Les quatre scores de bien-être ainsi que le score FACT-G global sont généralement standardisés de 0 à 100 dans le but de faciliter la comparaison avec d'autres questionnaires.

Afin d'évaluer les symptômes et modalités de traitement propres à chaque localisation cancéreuse, le choix du groupe FACT a été d'ajouter une sous-échelle spécifique de la localisation concernée au sein du questionnaire FACT-G. A titre d'exemple, le questionnaire FACT-L spécifique du cancer pulmonaire contient les 27 items du FACT-G et 9 items supplémentaires permettant d'évaluer une sous-échelle spécifique du cancer du poumon. D'autre part, un score global FACT-L est estimé et correspond à la moyenne des sous-scores obtenus pour chaque dimension et pour la sous-échelle additionnelle spécifique au cancer du poumon. Enfin, selon les questionnaires, des scores indices sont également proposés. Concernant le questionnaire FACT-L, l'indice « Trial Outcome Index » (TOI) correspond à un score résumé du bien-être physique et fonctionnel et de la sous-échelle spécifique au cancer du poumon. Ces scores résumés (score global FACT-G, score global FACT-L et indice TOI) peuvent alors être utilisés dans les analyses de données de QdV.

### **c. Administration des questionnaires**

Différents mode d'administration des questionnaires existent, comme le mode « papier-stylo », par téléphone ou lors d'un entretien avec un attaché de recherche clinique. Le gold standard pour l'évaluation et le mode de remplissage des questionnaires reste à ce jour le mode « papier-stylo ». Le mode d'administration des questionnaires peut influencer les résultats et le mode « papier-stylo » semble ainsi le moins susceptible de présenter des biais de mesure comme le biais de désirabilité sociale (50,51). Des outils informatisés sont également de plus

en plus développés. Le groupe EORTC a par exemple développé un programme informatisé pour l'évaluation et l'analyse de la QdV (52).

Une personne responsable de l'administration des questionnaires doit être clairement identifiée dans chaque centre. La plupart du temps il s'agit d'un attaché de recherche clinique. Cette personne doit savoir quand le patient arrive, s'assurer que le patient reçoive bien le questionnaire dans un endroit calme avant l'annonce d'un diagnostic ou une procédure thérapeutique et implémente les données de suivi au cas où le patient ne soit pas disponible. Il est important que l'administration des questionnaires se fasse les jours de traitement ou de consultation de suivi afin de limiter les données manquantes. Le lieu et la fenêtre temporelle pour le remplissage peuvent également influencer le choix des réponses du patient. En effet, il a été démontré que si un patient complète le questionnaire à domicile il ne donnera pas nécessairement les mêmes réponses que s'il était à l'hôpital (53). De plus, le remplissage du questionnaire pourra être différent selon que le patient le remplisse avant ou après un entretien avec son médecin (54). L'aide de l'attaché de recherche clinique peut être utile si le patient présente des difficultés de lecture ou d'écriture. La procédure d'administration des questionnaires doit être décrite dans le protocole.

#### **d. Temps d'évaluation de la qualité de vie**

Dans les essais cliniques en général, au moins trois évaluations de la QdV sont réalisées et requises : à l'inclusion dans l'étude (avant le début de l'administration du traitement), pendant le traitement et à la fin de l'étude après arrêt du traitement. Le nombre et le moment de passation des questionnaires dépendent à la fois des objectifs de l'étude et de la faisabilité. Néanmoins, un minimum de deux mesures est nécessaire : à l'inclusion et à la fin de l'étude ou après un arrêt prématuré du traitement (55).

Il est essentiel que la première évaluation de la QdV ait lieu non seulement avant le début du traitement, mais également avant la randomisation du traitement. En effet, la QdV étant un critère de jugement subjectif, la connaissance de l'attribution du traitement par le patient pourrait influencer son évaluation de la QdV (56).

Durant le traitement, deux types de schémas d'évaluations peuvent être réalisés. Une évaluation « Event-driven design » sera réalisée si les schémas d'administration de traitements diffèrent entre les deux bras considérés. L'évaluation est généralement planifiée juste avant une intervention telle que la chirurgie ou le début de la radiothérapie. L'objectif de

de ces évaluations est d'étudier l'effet du traitement sur la QdV à court terme. En revanche, lorsque l'objectif est d'étudier la QdV à long terme et que les deux types de prise en charge respectent le même schéma, une évaluation de type « Time-driven design » est réalisée. Dans ce type de design, les temps d'évaluation sont basés sur des intervalles de temps régulièrement espacés. Par exemple, tous les mois pendant trois mois, puis tous les trois mois pendant un an. Certaines études peuvent également être basées sur un mélange de ces deux types de schémas d'évaluation.

La fréquence des évaluations dépend de la localisation cancéreuse, de la situation thérapeutique ainsi que de la probabilité d'observer un changement de la QdV durant cette période. Les évaluations doivent être assez nombreuses pour pouvoir capturer un changement significatif d'un point de vue clinique. En revanche, des évaluations en trop grand nombre ou trop rapprochées dans le temps risqueraient également d'ennuyer ou de lasser le patient. Si un changement de QdV est attendu dans les premiers mois de l'étude par exemple, alors des évaluations plus fréquentes de la QdV doivent avoir lieu durant cette période.

L'évaluation de la QdV est généralement planifiée jusqu'à la progression de la maladie et/ou l'arrêt du traitement. Pour des aspects pratiques et scientifiques, des recommandations doivent être développées concernant le suivi des patients qui n'ont pas suivi le protocole du traitement. En effet, un bras de traitement avec un fort taux de sorties d'étude prématurées pourrait paraître bénéfique puisque seuls les patients en meilleure santé seraient maintenus dans l'étude.

L'objectif de ces évaluations de la QdV au cours du temps est d'étudier l'impact du traitement et de la progression de la maladie sur la QdV du patient. Une étude longitudinale de ces données de QdV est donc requise.

#### **e. Les dimensions cibles**

Une fois que le choix du (ou des) questionnaire(s) a été fixé, il est nécessaire de cibler les dimensions d'intérêt premier (entre deux et cinq dimensions) afin de prévenir l'inflation du risque de première espèce. Ces dimensions doivent être clairement identifiées dans le protocole de l'étude. Les autres dimensions seront étudiées à titre exploratoire ou comme objectif secondaire.

Les dimensions prioritaires doivent être, dans la mesure du possible, des dimensions multi-items. En effet, l'évaluation des domaines de QdV d'intérêt premier sera d'autant plus précise que le nombre d'items est important.

Afin de prévenir l'inflation du risque de première espèce dû à la multiplicité des analyses, il est recommandé de faire un ajustement du risque de première espèce, en utilisant par exemple un ajustement de Bonferroni, consistant à diviser le risque de première espèce par le nombre de dimensions de QdV retenues. Cet ajustement peut être évité si la QdV est un critère secondaire de l'étude.

Lorsque plusieurs dimensions de QdV sont étudiées, les règles décisionnelles des critères de jugements multiples s'appliquent. Selon les situations, on pourra considérer que la mise en évidence d'au moins un résultat significatif est suffisante (« single sufficient rule »). Inversement, on pourra considérer que l'étude est significative si toutes les dimensions sont significatives.

Une procédure des tests fermés (« close testing ») peut également être réalisée. Dans ce cas, un test global est tout d'abord réalisé sur l'ensemble des  $n$  dimensions de QdV au niveau de signification  $\alpha$  sans ajustement. Si le test s'avère significatif, tous les tests globaux intégrant  $n-1$  dimensions de QdV seront réalisés. A l'étape suivante, les tests comprenant  $n-2$  dimensions de QdV sont réalisés seulement si tous les tests incluant  $n-1$  critères de jugement sont significatifs, et ainsi de suite jusqu'à réalisation des tests n'intégrant qu'une seule dimension de QdV.

#### **f. La différence minimale cliniquement importante**

La différence minimale cliniquement importante (DMCI) représente le plus petit changement de score qui est cliniquement pertinent. Déterminer la DMCI pour l'interprétation des scores de QdV est indispensable tant pour pouvoir évaluer l'efficacité d'une prise en charge que pour déterminer le nombre de sujets nécessaires pour la réalisation d'un essai clinique.

Il n'existe pas à ce jour de méthode standard pour déterminer la DMCI (57). Les méthodes proposées sont généralement regroupées en deux catégories : les méthodes basées sur l'ancre (« anchor-based method ») et les méthodes basées sur la distribution (« distribution-based method »).

Les méthodes basées sur l'ancre se focalisent sur la relation entre les scores de QdV étudiés et un critère externe ayant un sens clinique. Ce critère externe correspond à l'ancre. Cette ancre peut être un indicateur de la progression de la maladie ou de la réponse clinique, mais également une évaluation effectuée par le patient lui-même du changement de la QdV globale. Ce dernier type de critère externe est celui le plus souvent utilisé comme ancre pour déterminer la DMCI (58,59). A titre d'exemple, la question de la transition de Jaeschke est régulièrement utilisée comme ancre (60). Cette question est généralement formulée de la façon suivante :

« Durant les trois derniers mois, considérez-vous que votre niveau de qualité de vie :

- N'a pas changé globalement
- S'est détérioré : un peu, assez, beaucoup
- S'est amélioré : un peu, assez, beaucoup ».

Plusieurs ancres peuvent également être utilisées dans la même étude afin de tester la cohérence entre les résultats (61).

Les méthodes basées sur la distribution considèrent que la DMCI peut être estimée en se basant sur la distribution des scores observés sur un échantillon représentatif de la population considérée. Certains chercheurs suggèrent que la moitié de l'écart type (62) ou l'erreur standard de mesure (« Standardized Response Mean ») (63,64) peuvent être considérés comme la DMCI. Cependant, cette différence observée est sûrement cliniquement importante mais ne correspond pas nécessairement à la différence minimale cliniquement importante. Il est nécessaire de résumer ces informations en termes d'« Effect Size » pour pouvoir apprécier le degré de changement. De plus, ces méthodes basées sur la distribution ne fournissent pas d'information directe sur la DMCI, seules les méthodes basées sur l'ancre permettent de mesurer directement la DMCI.

Concernant le questionnaire EORTC QLQ-C30 une DMCI de 5 ou 10 points est couramment utilisée (65,66) et 7 points pour le FACT-G (67,68). Maringwa et al. ont mené une étude pour déterminer la DMCI pour les dimensions du QLQ-C30 dans les cancers pulmonaires. Une DMCI de 5 à 10 points semble la plus pertinente dans les cancer pulmonaires (69).

Cependant, comme la DMCI dépend des caractéristiques de la population et du contexte clinique de l'étude, il n'existe pas nécessairement une seule valeur de DMCI pour un même questionnaire pour toutes les populations considérées (59). De plus, la DMCI pour un même questionnaire peut dépendre du sens du changement considéré, i.e. peut varier selon que l'on considère une amélioration du score ou une détérioration (70). Généralement, elle est plus

importante pour la détérioration que pour l'amélioration (71). Une autre problématique de la DMCI concerne l'influence du niveau de QdV des patients à l'inclusion (72). Certaines études ont ainsi proposé de définir la DMCI en fonction du niveau du score à l'inclusion et non comme une valeur unique (58,59). En effet, une valeur de DMCI unique suppose que le score a des propriétés de mesures d'intervalles, i.e. qu'une augmentation ou une diminution de  $\Delta$  points du score du patient a la même signification au regard du vrai niveau de QdV du patient quel que soit le niveau du score initialement observé. Les scores obtenus d'après les réponses aux items respectent rarement cette propriété. Ces scores correspondent davantage à des mesures ordinales, ayant des propriétés d'ordonnement, i.e. qu'on peut classer les patients selon les valeurs de scores observées : par exemple, un patient avec un score faible présente un moins bon niveau de QdV qu'un patient avec un score élevé. Ainsi, une différence de  $\Delta$  points sur une échelle de 0 à 100 n'aura peut-être pas nécessairement la même signification pour un patient avec un score faible correspondant à un bas niveau de QdV que pour un patient avec un score élevé et correspondant à un haut niveau de QdV. Enfin, la problématique des effets planchers et des effets plafonds compromet également l'observation d'un changement important même si ce changement a lieu.

Les DMCI retenues généralement pour les questionnaires EORTC et FACT sont donc très approximatives et susceptibles de varier entre les populations et les questionnaires utilisés. Cependant, ces valeurs sont généralement celles utilisées dans les études longitudinales pour une question de simplicité.

## **g. Analyse statistique longitudinale de la qualité de vie**

### **i. Population d'analyse**

Avant de réaliser l'analyse longitudinale des données de QdV, il est nécessaire de définir la population d'analyse. Afin de pouvoir interpréter au mieux les résultats et d'éviter tout biais de sélection de patients, l'idéal serait que la population d'analyse de la QdV soit la même que celle du critère de jugement principal (45). Généralement, tous les patients randomisés, quel que soit le respect des critères d'éligibilité sont analysés : il s'agit de l'analyse en intention de traiter. Dans le cadre de l'étude de la QdV, la présence d'au moins une mesure de QdV par patient est nécessaire pour pouvoir réaliser l'analyse : il s'agit alors d'une analyse en intention de traiter modifiée. Ainsi, l'analyse peut être réalisée en sélectionnant les patients ayant au



moins la mesure initiale de disponible. Dans tous les cas, cette population doit être clairement définie en amont de l'analyse.

## ii. Méthodes d'analyse longitudinale

Deux principales techniques d'analyse longitudinale sont utilisées : les modèles mixtes d'analyse de variance sur mesures répétées et temps jusqu'à détérioration d'un score de QdV (TJD).

### • Les modèles mixtes d'analyse de variance sur mesures répétées

Les modèles mixtes d'analyse de variance sur mesures répétées sont des modèles complexes qui permettent de modéliser des données telles que, pour chaque individu considéré, on dispose d'observations à différents instants, autrement dit répétées dans le temps. Elle consiste à appliquer un modèle linéaire à effets mixte sur le score observé et calculé à chaque temps de mesure. Dans le cadre d'une analyse longitudinale de QdV, ces modèles permettent d'estimer l'effet du temps (c'est-à-dire les changements de score au cours du temps), l'effet groupe (i.e. des différences de scores entre les bras de traitement à l'inclusion) et une interaction entre le bras de traitement et le temps (reflétant une évolution différente du score de QdV selon le bras de traitement au cours du temps comparativement au bras de référence). Ils contiennent des effets fixes comme le traitement et des effets aléatoires comme l'effet patient en prenant en compte la corrélation entre les mesures faites pour un même patient à différents moments. Pour cela, le type de matrice de covariance entre les mesures de QdV doit être précisé. Les structures de matrice les plus utilisées sont les non-structurées, auto-régressives d'ordre 1 et à symétrie composée. Les critères d'information Akaike Information Criteria (AIC) et Bayesian Information Criteria (BIC) sont utilisés pour sélectionner la meilleure structure.

Le modèle peut s'écrire :

$$Y_i = X_i\beta + Z_i\gamma_i + \epsilon_i$$

Où :

- $Y_i$  correspond au vecteur des scores du patient  $i$  ;
- $\beta$  est un vecteur d'effets fixes ;
- $Z_i$  est une sous-matrice de  $X_i$  pour les effets aléatoires ;

- $\gamma_i$  est un vecteur d'effets aléatoires de distribution normale, de moyenne 0 et de matrice de covariance  $\Sigma$  ;
- et  $\epsilon_i$  correspond à l'erreur aléatoire et suit une loi normale de moyenne 0 et d'écart type  $\zeta$ .

Une limite de ce modèle est qu'il requiert généralement la normalité du score. De plus, les résultats sont peu familiers pour les cliniciens.

- **Le temps jusqu'à détérioration d'un score de qualité de vie relative à la santé**

Depuis quelques années, une méthode d'analyse de survie basée sur le temps jusqu'à détérioration (TJD) d'un score de QdV est régulièrement utilisée pour l'analyse longitudinale de la QdV, en particulier dans les essais cliniques de phase III (27,43,73,74). Cette méthode a été proposée pour la première fois en 2002 par Award et al (75).

Le TJD nécessite une définition de l'évènement, i.e. de la détérioration. Cette définition dépend du choix du score de référence, de la DMCI considérée, de la considération des scores manquants et de l'intégration ou non du décès dans la définition de l'évènement.

La définition la plus intuitive du TJD est le délai depuis la date d'inclusion ou de randomisation jusqu'à l'apparition d'une première détérioration significative du niveau de QdV par rapport au niveau à l'inclusion (76). Les patients ne présentant pas de détérioration sont censurés à la date de dernière mesure de la QdV ou de dernière nouvelle. Dans cette définition, la détérioration observée est un état transitoire et non nécessairement définitif.

Le temps jusqu'à détérioration définitive (TJDD) d'un score de QdV a été défini en situation avancée comme l'intervalle de temps depuis la date d'inclusion ou de randomisation et l'apparition d'une première détérioration significative par rapport au score obtenu à l'inclusion, sans amélioration ultérieure significative par rapport au score à l'inclusion, ou le décès. Un patient présentant une détérioration significative du niveau de QdV suivi de données manquantes monotones dû à la sortie de l'étude du patient (drop-out) ou au décès est considéré en détérioration définitive. Les patients ne présentant pas de détérioration et ceux présentant une détérioration mais suivie d'une amélioration significative par rapport au score à l'inclusion sont censurés au moment du remplissage du dernier questionnaire de QdV.

Les données manquantes intermittentes ne sont généralement pas prises en compte dans ces modèles, considérant que le niveau de QdV reste constant entre deux mesures de QdV consécutives.

L'analyse du TJD d'un score de QdV est une approche de type analyse de survie. Il peut être défini comme l'intervalle de temps entre la date de randomisation et l'observation d'une première diminution d'une valeur seuil de la DMCI d'un score de QdV par rapport au score à l'inclusion. De multiples définitions peuvent être proposées en fonction : de la prise en compte du décès comme événement, du choix du score de référence (score à l'inclusion, score précédent, meilleur score précédent), de la valeur seuil de la DMCI (77).

L'estimation du TJD se fait généralement selon la méthode de Kaplan-Meier. Une comparaison entre les deux bras de traitement peut être réalisée selon le test du log-rank et un modèle de Cox multivarié peut être réalisé afin d'estimer la taille de l'effet traitement (hazard ratio).

#### **h. Valeur pronostique de la qualité de vie relative à la santé sur la survie globale**

La valeur pronostique de la QdV sur la survie globale a été démontrée dans de nombreuses situations, en particulier en phase métastatique (78–83). Quinten et al. ont conduit une méta-analyse sur données individuelles de 10000 patients atteints de 11 type de cancers différents. Ils ont démontré la valeur pronostique de la QdV sur la survie globale.

Concernant les CBNPC, plusieurs études ont démontré le caractère pronostique de la QdV (79,84–86). Par exemple, Sloan et al. ont mené une étude prospective sur 2442 patients atteints de CBNPC stade I-IV qui ont complété un item du questionnaire Lung Cancer Symptom Scale dans les 6 mois après le diagnostic. Ils ont démontré que la QdV était associé à la SG (hazard ratio [HR]: 1.55;  $P < 0.001$ ). Néanmoins aucune étude n'a été spécifiquement conduite sur les sujets âgés.

## *Objectifs*

---

« Pour expliquer un brin de paille, il faut démontrer tout l'univers »

*Rémy de Gourmont*

La SG est considérée comme le critère de jugement principal de référence et le plus pertinent dans les essais cliniques en oncologie. La US Food and Drug Administration (FDA) estime la SG comme étant une mesure unanimement admise pour évaluer le bénéfice direct du traitement et doit être le critère de jugement privilégié pour les essais de phase III. Dans son rapport concernant spécifiquement les CBNPC (Clinical Trial Endpoints for the Approval of Non-Small Cell Lung Cancer Drugs and Biologics Guidance for Industry) la FDA précise: “OS is an optimal endpoint because the measurement is accurate, is observed on a daily basis, and provides direct evidence of clinical benefit to the patient. Regular approval was granted on the basis of a significant improvement in OS”.

Cependant les critères de jugement intermédiaires centrés sur la tumeur sont fréquemment utilisés par les oncologues car ils sont disponibles plus précocement que la SG, moins influencés par les causes compétitives de décès et ne dépendent pas des traitements administrés après la progression. Néanmoins, ils présentent d’importantes limites, notamment leur hétérogénéité de définition et leur absence de validation comme critère de substitution de la SG.

De ce fait, la QdV est reconnue par l’ASCO et la FDA comme un critère de jugement des essais cliniques en oncologie qui reflète un réel bénéfice clinique pour le patient. La FDA souligne l’importance de la mesure de la QdV dans les essais cliniques des CBNPC : “Patient-reported outcome (PRO) measures of tumor-related symptoms and functioning can provide direct evidence of treatment benefit if demonstrated to be well-defined and reliable assessments of a clinically meaningful concept or set of concepts, and if evaluated in well-conducted, placebo-controlled or double-blinded, randomized trials”. Néanmoins la QdV est un phénomène complexe et de nombreuses méthodes d’analyse longitudinale existent mais à ce jour aucun standard n’a pu être établi.

Dans ce contexte, les objectifs de ce travail ont été :

- Etudier les différents types et définitions des critères de jugement utilisés dans les essais cliniques de phase III des CBNPC avancés ;
- Etudier les critères de jugement intermédiaires des essais cliniques de phase III des cancers pulmonaires validés comme critères de substitution ;
- Etudier la méthodologie d’analyse de la QdV dans les essais cliniques de phase III des CBNPC avancés ;
- Proposer et réaliser l’analyse de la QdVselon la technique du TJD de l’essai clinique IFCT-0501 ;

- Etudier la valeur pronostique de la QdV sur la SG dans l'essai clinique IFCT-0501 ;
- Discuter les problèmes méthodologiques soulevés par un design d'essai clinique avec la QdV comme co-critère de jugement principal.

## *Articles*

---

« Le chemin est long du projet à la chose.»

*Molière*

## Les critères de jugement dans les essais cliniques en cancérologie

### Résumé

Les critères de jugement sont des mesures cliniques et biologiques pour l'évaluation et le développement des thérapeutiques. En cancérologie, ils peuvent être classés en deux catégories principales : les « critères de jugement centrés sur le patient » qui comprennent la survie globale (SG) et la qualité de vie relative à la santé (QdV) et les « critères de jugement centrés sur la tumeur » comme la survie sans progression. Les critères de substitution sont des critères de jugement principalement centrés sur la tumeur dont l'objectif est de se substituer aux critères de jugement centrés sur le patient notamment la SG. Le choix des critères de jugement dans les essais cliniques en oncologie est une problématique majeure pour apprécier le réel bénéfice du traitement pour le patient. La publication du guide de bonnes pratiques CONSORT conseille de reporter précisément les définitions des critères de jugement primaires et secondaires. La SG est le critère de référence mais avec l'augmentation du nombre de traitements efficaces disponibles dans beaucoup de cancers, il est nécessaire d'inclure un nombre plus important de patients et de les suivre plus longtemps ce qui augmente le coût des essais cliniques. Ainsi, les critères de jugement dit intermédiaires centrés sur la tumeur qui sont évalués plus précocement et utilisés comme critères substitutifs de la survie globale sont de plus en plus utilisés mais la plupart d'entre eux n'ont pas de définitions standardisées ce qui limite la comparaison des résultats de différents essais cliniques et ne sont pas validés comme critères de substitution selon la méthodologie de référence. De plus, la variabilité de leur définition peut fortement impacter les conclusions des essais en influençant l'estimation et la puissance statistique. Dans ce contexte, la QdV constitue un critère de jugement utile pour les essais assurant une évaluation du bénéfice du traitement tant d'un point de vue du patient que d'un point de vue de la santé publique. Les recherches en méthodologie doivent se poursuivre pour développer des définitions consensuelles des critères de jugement des essais cliniques en cancérologie et pour proposer une analyse longitudinale standardisée des données de QdV. Dans cet article nous discutons les problématiques soulevées par les critères de jugement dans les essais cliniques en cancérologie.





Available online at  
**ScienceDirect**  
[www.sciencedirect.com](http://www.sciencedirect.com)

Elsevier Masson France  
**EM|consulte**  
[www.em-consulte.com/en](http://www.em-consulte.com/en)



REVIEW

## Endpoints in cancer clinical trials



F. Fiteni<sup>a</sup>, V. Westeel<sup>c</sup>, X. Pivot<sup>a</sup>, C. Borg<sup>a</sup>,  
 D. Vernerey<sup>b</sup>, F. Bonnetain<sup>b,\*</sup>

<sup>a</sup> *University Hospital of Besançon, Department of Medical Oncology, 3, boulevard Fleming, 25000 Besançon, France*

<sup>b</sup> *Methodology and quality of Life in Oncology unit (EA 3181) & quality of Life and cancer clinical research platform, CHU de Besançon, 2, place Saint-Jacques, 25000 Besançon, France*

<sup>c</sup> *University Hospital of Besançon, Chest disease Department, Besançon, France*

Available online 14 January 2014

### KEYWORDS

Endpoint definitions;  
 Cancer clinical trials;  
 Overall survival;  
 Surrogate endpoints;  
 Health-related  
 Quality of life;  
 Disease-free survival;  
 Progression-free  
 survival;  
 Methodological  
 researches

**Summary** Endpoints are measurable clinical and biological findings that are used for the development and assessment of treatment options. In the treatment of cancer, endpoints can be classified into two categories: “patient-centered clinical endpoints” including overall survival (OS) and health-related quality of life (QoL), and “tumor-centered clinical endpoints” such as progression-free survival. Surrogate endpoints are tumor-centered clinical endpoints that can be used as substitutes for patient-centered clinical endpoints, particularly OS. The choice of endpoints in oncology trials is a major problem. The published Consolidated Standards of Reporting Trials (CONSORT) best-practice guidelines encourage the reporting of clearly defined primary and secondary outcome measures. OS is the gold standard of endpoints but as increasing numbers of effective salvage treatments become available for many types of cancer, much larger numbers of patients are included; this requires a longer follow-up period and increases the cost of clinical trials. Thus, tumor-centered clinical endpoints that can be assessed earlier and used as surrogates for overall survival are increasingly studied, but most of them currently lack standardized definitions to enable cross comparison of results among different clinical trials and they have not been validated as surrogate endpoints. In addition, the variability of their definition can strongly impact the trial’s conclusions by affecting both statistical power and estimation. In this context, QoL constitutes an available and useful surrogate endpoint for trials to ensure treatment benefit from both the patient and public health points of view. Methodological research should be pursued to develop standard outcome definitions for use in cancer clinical trials and to define a standardized longitudinal analysis of QoL data.

© 2013 Elsevier Masson SAS. All rights reserved.

### Introduction

Primary or secondary endpoints in oncologic clinical trials can be classified into two main categories: “patient-centered clinical endpoints” including overall survival (OS) and health-related quality of life (QoL), and “tumor-centered clinical endpoints” such as progression-free survival [1–3]. The use of OS as a primary endpoint is the reference

\* Corresponding author.  
 E-mail address: [franck.bonnetain@univ-fcomte.fr](mailto:franck.bonnetain@univ-fcomte.fr) (F. Bonnetain).

standard to demonstrate patient clinical benefit, but the increasing number of effective salvage treatments for many types of cancer (i.e., second-line treatments) has resulted in the need for a larger number of patients to be included and/or the need of a more prolonged observation period to attain sufficient data that can achieve planned statistical power; this increases the cost of clinical trials and requires a longer duration to obtain results. Consequently "tumor-centered clinical endpoints" such as progression-free survival are often used because they can be assessed earlier (i.e., intermediate endpoints). However, they do not reflect a direct clinical benefit for the patient, there is a lack of consistency in their definitions, and they have not been systematically validated as surrogate endpoints for OS. In this context, health-related quality of life (QoL) constitutes an alternative endpoint that allows earlier assessment of direct clinical benefit for the patient.

Here, we focus on the different problems raised by the choice of primary endpoint in cancer clinical trials.

### Patient-centered clinical endpoints

Patient-centered clinical endpoints are variables that reflect a patient's feeling of well-being or his survival [1]. They assess the effects of a therapeutic intervention for the patients in clinical trials. In oncologic clinical trials, such endpoints include OS and QoL. They assess a direct patient clinical benefit [2,3].

### Overall survival

OS is considered as the gold standard and the most clinically relevant primary endpoint in oncology clinical trials. The US Food and Drug Administration (FDA) has considered OS as a universally accepted measure of direct treatment benefit and the preferred endpoint of Phase III clinical trials.

OS is defined as the time from randomization or initiation of treatment to patient death, irrespective of cause. Patients who are alive or lost to follow-up at the cut-off date are excluded.

Another endpoint is cancer-specific survival which is defined as the time from randomization or treatment initiation to patient death caused by the index cancer, whether due to the original tumor or to a second primary of the same cancer type. Deaths from other cancers, non-cancer-related deaths, treatment-related deaths, patients lost to follow-up and those alive at the cut-off date are excluded. The difficulty lies in identifying the specific cause of death. Moreover, this is a composite endpoint and therefore somewhat difficult to define.

However, with the increasing number of effective salvage treatments available for many types of cancer, much longer follow-up is necessary to observe the number of deaths required to achieve significant statistical power and then to establish treatments that may improve OS. In this context, we could be faced with the risk of a larger number of patients lost to follow-up (information bias). Moreover, for good-prognosis cancers, the expected difference between the two arms of treatment is often small thereby increasing the number of patient deaths needed in order to define a statistically significant difference in OS. The prolonged follow-up and the higher number of patients included increase the cost of clinical trials. Moreover, given the availability of other effective therapies administered after participation in the trial, interpretation of results can

be confounded by the effects of post-trial therapy. Current management of metastatic breast cancer illustrates this problem; it becomes increasingly difficult to demonstrate an OS benefit for first-line therapy, unless post-trial administration of other effective therapies is controlled for [33]. However, taking into account randomization, we could also speculate that administration of standard-of-care subsequent second- and third-line therapies will result in equal distribution between treatment arms, thereby avoiding confounding impact on OS.

All these reasons support the need for surrogate clinical endpoints that could be assessed earlier in the clinical course.

### Health-related Quality of life (QoL)

In cancer research, the number of studies that incorporate QoL assessment has been rising over the last decade. Two events have combined to promote QoL as an endpoint: the creation of the World Health Organization (WHO) and the establishment of a multidimensional definition of health that encompasses physical, emotional, and social well-being [4]. QoL could be considered as a patient-defined measure using the WHO health definition of QoL. There is general agreement concerning the multidimensional concept of QoL taking into account levels of physical, mental, social, and patient satisfaction with treatment as well as treatment outcome [5,6]. QoL reflects direct clinical benefit of treatments for the patient. Many validated self-completion QoL questionnaires for cancer patients are available: e.g. the European Organization for Research and Treatment of Cancer QoL Questionnaire (EORTC QLQ-C30), the Functional Assessment of Cancer Therapy (FACT), the Rotterdam Symptom Checklist (RSCL), and the Functional Living Index-Cancer (FLIC). The QLQ-C30 is the most widely used QoL measurement tool in Phase III cancer clinical trials. It is composed of 30 items that generate 15 scores: five functional scales, three symptom scales, a global health status/QoL scale, and six single items. It can be supplemented by different modules for specific cancer types. EORTC recommends a minimum of three measurements during clinical trials: before, during, and after treatment [7]. A more intensive longitudinal assessment of QoL is generally encouraged to ensure the capture of meaningful clinically-important differences.

The FDA considers QoL as an endpoint in assessing clinical benefit. Consequently QoL has emerged as a primary endpoint candidate in oncology clinical trials, offering two advantages: a shorter study duration and an ability to assess clinical patient benefit. Opponents of QoL as a primary endpoint note its limitations due to its subjective nature, incomplete data, and the lack of standardization for analyzing and reporting QoL results. Moreover, it is difficult to provide analysis of clinically interpretable QoL results.

The evidence that QoL analysis presents methodological and statistical difficulties due to the type of data generated and the multidimensional nature of the measurement tools should be addressed to ensure comparability of trial results. One of the major concerns is missing data. Indeed, in longitudinal studies, patient assessments can be missed at specified time points because the patients miss visits or fail to complete prescribed questionnaires [8–10]. Consequently, potential biases limit confidence in QoL results, even though many statistical solutions have been robustly tested and proposed, such as multiple imputation and logistical prevention of missing QoL questionnaires. There is also

a need to define a standardized analysis of longitudinal QoL data and to develop ways of presenting results that are clinically meaningful in order to assist clinical decision-making.

## Tumor-centered clinical endpoints

Tumor-centered clinical endpoints include biological markers. In most oncologic clinical trials, these endpoints include laboratory or histology markers to define response to a therapeutic intervention [1]. For instance, tumor response (RECIST), circulating tumor cells, disease-free survival (DFS) and progression-free survival (PFS) are tumor-centered clinical endpoints. Nevertheless, these criteria might not directly reflect clinical benefit to the patient. Conclusions of trials using such endpoints remain questionable and complementary studies to assess patients' QoL are often requested by regulatory agencies. Their use as surrogate endpoints for patient-centered clinical endpoints must be validated.

## Surrogate endpoints

A surrogate endpoint is a tumor-centered clinical endpoint used as a substitute for a patient-centered clinical endpoint. In oncology, endpoints that could be considered for use as surrogates for OS but assessable earlier in the patient's course have been increasingly studied. Statistical validation of a surrogate endpoint is a difficult but necessary process. For example, Shindoh et al. [11] demonstrated that optimal morphologic response (tumor shrinkage) in response to pre-operative chemotherapy before undertaking resection of hepatic colorectal metastases is correlated with improved OS; they concluded that optimal morphologic response can be considered as a surrogate endpoint of OS. Nevertheless, a simple correlation between the proposed surrogate and OS is not sufficient since we need to demonstrate that the treatment effect on the candidate surrogate endpoint is correlated to the treatment effect on OS and also that the surrogate endpoint correlates with OS.

To achieve these requirements, two main methods of validation have been developed: one allowing validation through a single clinical trial and the other requiring a meta-analytic approach.

Single trial validation includes two approaches: the use of Prentice criteria and the proportion of treatment effect attributable to a substitution criteria [12].

In 1989, Prentice explained that a surrogate endpoint should "yield unambiguous information about differential treatment effects on the true endpoint" and he defined a valid surrogate endpoint as 'a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint' [13]. Prentice proposed four operational criteria:

- treatment must have a significant effect upon the surrogate endpoint;
- treatment must have a significant effect upon the true endpoint;
- the surrogate endpoint must have a significant effect upon the true endpoint;
- the full effect of treatment upon the true endpoint must be mediated by the surrogate.

However, the demonstration of the last criterion is difficult because failure to reject the null hypothesis may be due to lack of power.

Freedman et al. [14] proposed the quantification of the proportion of the treatment effect explained by the surrogate. The quantification of proportion is expressed as:  $PE = 1 - (\beta S / \beta)$  where  $\beta$  and  $\beta S$  are the estimates of the effect of the treatment on the true endpoint without and with adjustment for the surrogate endpoint. The most serious difficulty is the possibility for  $PE$  to be  $> 1$  or to be negative, which can hardly be justified for a proportion [14,15].

The meta-analytic method is now the gold standard for validation of a surrogate endpoint. Based on the results of previously completed trials [8,13] and using individual data, this approach jointly estimates:

- the correlation between the candidate surrogate and the final endpoints;
- the correlation between the treatment effect on the candidate surrogate and its effect on OS.

Thus, the surrogate endpoint must first be a prognostic marker, without reference to any specific intervention. This aspect of surrogacy is called 'individual-level' surrogacy, which means that for an individual patient, the surrogate endpoint must correlate with the final endpoint of interest, such as OS. Secondly, the effect of treatment on the candidate surrogate endpoint must be correlated with the effect of treatment on the true clinical endpoint. This aspect of surrogacy is called 'trial-level' surrogacy since it must be demonstrated for a group of patients in a clinical trial [16].

Nevertheless, a major difficulty for the validation of surrogate endpoints arises from the fact that they must be validated with respect to a specific therapeutic class, for a specific disease, at a specific stage (i.e., 5-FU chemotherapy for first line treatment in metastatic colorectal cancer). It is uncertain if the same surrogacy relationship will be applicable for a different treatment or at another stage of the disease [16]. Then, since trials must be performed in order to validate a surrogate for a new therapeutic class, clinical benefit in trials using such endpoints remains uncertain.

Currently, in cancer clinical trials of 5FU-based chemotherapy for advanced colorectal cancer, only disease-free survival has been validated as a surrogate for OS [17]. In T3 and T4 rectal cancers, while pathological complete response to neo-adjuvant therapy and R0 resection are important endpoints for local control, they have not been validated as surrogate endpoints of OS [18]. The ARCAD Clinical Trials Program is a worldwide collaboration of clinicians, statisticians and scientists specializing in gastrointestinal malignancies who have collaboratively constructed the ARCAD Advanced Colorectal Cancer Database, integrating patient-level data from most recent major trials in advanced colorectal cancer into a single resource [19–21]. New results regarding surrogate endpoints from the ARCAD database including biotherapies will be presented in 2013.

## Tumor-centered endpoints: the example of disease-free survival (DFS) and progression-free survival (PFS)

While overall survival is the most clinically relevant endpoint in cancer clinical trials, there is a need for surrogate endpoints that could be assessed earlier. Tumor-centered endpoints are composite endpoints, based on tumor assessment, and they combine different events such as local and

distant progression, local and distant recurrence, development of metachronous cancer, death, or severe toxicity. PFS is generally the time between random treatment assignment and tumor progression or death resulting from any cause in the metastatic stage. DFS is generally defined as the time from randomization until tumor recurrence or any-cause death after treatments given with curative intent. These endpoints are frequently used by oncologists as primary outcomes in oncologic clinical trials because they are available earlier than OS, less influenced than OS by competing causes of death, and not influenced by treatments administered after progression. Nevertheless, they suffer from important limitations:

- there are a variety of composite endpoints used as primary outcomes in oncologic clinical trials. Time to tumor progression (TTP) differs from PFS in that the event of interest is only disease progression, while patients who die of other causes are not included. PFS and TTP have often been used interchangeably in the recent literature. Consequently cross comparison of results from different clinical trials is often difficult;
- there is no international consensus standard for the definition of PFS and DFS; consequently the definition of the same "endpoint" is variable in different studies of the same disease. For example, the four randomized trials of adjuvant trastuzumab for HER-2 positive breast cancer illustrate the confusion that may arise from the use of heterogeneous definitions of DFS (Table 1) [22–25]. Moreover, Birgisson et al. [26] demonstrated that the inclusion of a second primary cancer other than the incident colorectal cancer as an event in the definition of disease-free survival (DFS) significantly impacted the results. The estimated DFS rate for patients with stage I–III disease was 62% after 5 years if this event was not counted as an event, compared with 58% if it was. The difference was larger for stage II (68 versus 60%) than for stage III (49 versus 47%). Another example is the PETACC 03 randomized study [27] where results were either significant or non-significant depending on whether or not second primary tumors were accounted for in the DFS definition [28];
- baseline tumor size measurement can be determined before treatment or after optimal tumor response. According to RECIST 1.1 criteria, tumor progression is defined as an increase of at least 20% in the sum of diameters of target lesions compared with the smallest sum on record (this includes the baseline sum if that is the smallest on record);
- tumor progression is variably defined in different trials (clinical or radiological progression (RECIST criteria) or laboratory measures of progression). This complicates comparisons of results between different trials. Moreover, symptomatic or non-radiologic progression is often subjective, typically lacks a standard definition and is a potential source of bias [29];
- disease progression is assessed on the basis of radiologic testing at scheduled times but the true moment of progression actually lies somewhere within the time interval between two radiologic studies; in addition, surveillance intervals vary widely between and even within trials [30];
- ascertaining radiological disease progression is subject to measurement error and bias;
- PFS has been validated as a surrogate endpoint only for 5-FU-based chemotherapy in advanced colorectal cancer [18]. Nevertheless, PFS is frequently used as the primary outcome in other types of cancer and many trials have demonstrated improved PFS with no improvement of OS.

For instance, the E2100 trial [34] studied bevacizumab plus weekly paclitaxel in metastatic breast cancer and showed a 6-month improvement in median PFS compared with paclitaxel alone. As a consequence, the FDA accelerated approval for its use. Subsequent double-blind studies with other chemotherapies confirmed more modest but still statistically significant improvements in PFS. None of these studies showed improvement in overall survival. In view of the absence of survival benefit, the modest PFS benefits noted in later studies together with toxic effects from all disease indications, the FDA took the unusual step of withdrawing approval for bevacizumab in the treatment of metastatic breast cancer. Conversely, the European Medicines Agency (EMA) approved the use of bevacizumab with paclitaxel and capecitabine [31]. This shows that the demonstration of clinical benefit in oncology trials using these endpoints has been increasingly questioned and will remain the subject of debate [32].

The variety of composite time-to-event tumor-centered endpoints and the variability of their definitions, particularly PFS and DFS, are recognized as a major methodological problem and efforts must be pursued to improve their reliability and reproducibility.

## Future perspectives

Given all the problems raised by the choice of endpoints in oncologic clinical trials, the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project has been developed to provide recommendations and to standardize definitions of time-to-event endpoints for each specific diseases and at each specific stage by use of a formal consensus methodology [29]. This is the preliminary step before assessing their surrogate capabilities (DATECAN 2 project). Nevertheless, when evaluating a new class of therapy, trials must be performed to validate a surrogate endpoint; therefore the clinical benefit of these trials using such endpoints remains uncertain.

Consequently, health-related quality of life (QoL) constitutes an alternative endpoint that allows earlier assessment of direct clinical benefit for the patient. However, there is a need to define a standardized longitudinal analysis of QoL data and a method to present results in a clinically meaningful way in order to assist clinicians in their decision-making. One pathway under study is the development of an endpoint called "Time to deterioration of QoL," defined as the time from inclusion in the study to the time of deterioration of the QoL score [9,10]. Several definitions of "Time to QoL deterioration" can be proposed depending on disease stage (adjuvant vs. metastatic chemotherapy); for this reason, practice guidelines similar to response evaluation criteria in solid tumors (RECIST) criteria for radiologic tumor evaluation should be defined, and standardized recommendations should be developed for application to QoL.

Finally, tumor-centered endpoints, for example, pathological complete response and local control in rectal cancers could be combined with QoL as co-primary endpoints to ensure a clinical benefit for patients [18]. However, methodological research must be pursued to clearly define decision-making rules for the design of such trials using two co-primary endpoints: one patient-centered and one tumor-centered.

Table 1 Example of inconsistent definitions of disease-free survival (DFS) in the four randomized trials of trastuzumab for HER-2 positive breast cancer in the adjuvant setting.

Trials	Events used to define DFS as primary time-to-event endpoint							
	Local/regional recurrence	Distant metastases	Second primary cancer	Contralateral invasive cancer	Death from any causes	Death before recurrence or a second primary cancer	Ipsilateral DCIS	Contralateral DCIS
HERA [24]	X	X	X	X		X		
NSABP B-31/ NCCTG N9831 [22]	X	X		X		X		X
BCIRG 006 [23]	X	X	X		X			
FinHer [25]		X		X				

DCIS: ductal carcinoma in situ.

## Conclusion

Overall survival remains the gold standard endpoint in oncologic clinical trials even though OS is difficult to measure. For that reason, surrogate endpoints of OS have been developed but their statistical validation is a difficult process. PFS and DFS are frequently used as primary endpoints but suffer from important limitations, particularly heterogeneity and a lack of validation as surrogates of OS. In this context, QoL could serve as an endpoint to judge treatment efficacy, particularly in advanced cancer, but this too needs methodological improvements and standardization. Tumor-centered endpoints could be combined with QoL as co-primary endpoints to ensure a clinical benefit for the patients [18].

## Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

## References

- [1] Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 2001;69:89–95.
- [2] Ohorodnyk P, Eisenhauer EA, Booth CM. Clinical benefit in oncology trials: is this a patient-centred or tumour-centred endpoint. *Eur J Cancer* 2009;45:2249–52.
- [3] Booth CM, Ohorodnyk P, Eisenhauer EA. Call for clarity in the reporting of benefit associated with anticancer therapies. *J Clin Oncol* 2009;27:213–4.
- [4] WHO. WHO constitution. Geneva: WHO; 1948.
- [5] Wood-Dauphinee S. Assessing quality of life in clinical research: from where have we come and where are we going? *J Clin Epidemiol* 1999;52:355–63.
- [6] Bonnetain F. Health related quality of life and endpoints in oncology. *Cancer Radiother* 2010;14:515–8.
- [7] Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993;85:365–76.
- [8] Post WJ, Buijs C, Stolk RP, et al. The analysis of longitudinal quality of life measures with informative drop-out: a pattern mixture approach. *Qual Life Res* 2010;19:137–48.
- [9] Hamidou Z, Dabakuyo TS, Mercier M, et al. Time to deterioration in quality of life score as a modality of longitudinal analysis in patients with breast cancer. *Oncologist* 2011;16:1458–68.
- [10] 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;46:2753–62.
- [11] Shindoh J, Loyer EM, Scott Kopetz, et al. Optimal morphologic response to preoperative chemotherapy: an alternate outcome end point before resection of hepatic colorectal metastases. *J Clin Oncol* 2012;30:4566–72.
- [12] Methy N, Bedenne L, Bonnetain F. Validation of surrogate endpoints in digestive oncology. *Bull Cancer* 2009;96:591–5.
- [13] Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 1989;8:431–40.
- [14] Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Stat Med* 1992;11:167–78.
- [15] Buyse M, Molenberghs G, Burzykowski T, et al. The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 2000;1:49–67.
- [16] Buyse M, Sargent DJ, Grothey A, et al. Biomarkers and surrogate end points the challenge of statistical validation. *Nat Rev Clin Oncol* 2010;7:309–17.
- [17] Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007;25:5218–24.
- [18] Bonnetain F, Bosset JF, Gerard JP, et al. What is the clinical benefit of preoperative chemoradiotherapy with 5FU/leucovorin for T3-4 rectal cancer in a pooled analysis of EORTC 22921 and FFCD 9203 trials: Surrogacy in question? *Eur J Cancer* 2012;48:1781–90.
- [19] de Gramont A, Haller DG, Sargent DJ, et al. Toward efficient trials in colorectal cancer: the ARCAD Clinical Trials Program. *J Clin Oncol* 2010;28:527–30.
- [20] <http://www.fondationarcad.org>
- [21] Buyse M, Sargent DJ, Goldberg R, de Gramont A, On behalf of the ARCAD Clinical Trials Program. The ARCAD advanced colorectal cancer database open for business. *Ann Oncol* 2012;23:281–2.
- [22] Romond EH, Perez EA, Bryant J, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* 2005;353:1673–84.
- [23] Slamon D, Eiermann W, Robert Nicholas, et al. Adjuvant trastuzumab in HER2-positive breast cancer. *N Engl J Med* 2011;365:1273–83.
- [24] Gianni L, Dafni U, Gelber RD, et al. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol* 2011;12:236–44.
- [25] Joensuu H, Bono P, Kataja V, et al. Fluorouracil, epirubicin, and cyclophosphamide with either docetaxel or vinorelbine, with or without trastuzumab, as adjuvant treatments of breast cancer: final results of the FinHer Trial. *J Clin Oncol* 2009;27:5685–92.
- [26] Birgisson H, Wallin U, Holmberg L, et al. Survival endpoints in colorectal cancer and the effect of second primary other cancer on disease free survival. *BMC Cancer* 2011;11:438.
- [27] Van Cutsem E, Labianca R, Bodoky G, et al. Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J Clin Oncol* 2009;27:3117–25.
- [28] Bellera C, Pulido M, Gorgou S, et al. 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* 2013;49:769–81.
- [29] Bhattacharya S, Gwen Fyfe G, Gray RJ, et al. Role of sensitivity analyses in assessing progression-free survival in late-stage oncology trials. *J Clin Oncol* 2009;27:5958–64.
- [30] Panageas KS, Ben-Porat L, Dickler MN, et al. When you look matters: the effect of assessment schedule on progression-free survival. *J Natl Cancer Inst* 2007;99:428–32.
- [31] Miles D. Bevacizumab in breast cancer: fundamental questions remain. *Lancet Oncol* 2013;14:99–101.
- [32] Booth CM, Eisenhauer EA. Progression-free survival: meaningful or simply measurable? *J Clin Oncol* 2012;30:1030–3.
- [33] Buyse M, Saad ED. Overall survival: patient outcome, therapeutic objective, clinical trial end point, or public health measure? *J Clin Oncol* 2012;30:1750–4.
- [34] Miller K, Wang M, Gralow J, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med* 2007;357:2666–76.

## **Les critères de jugement de type temps à évènement dans les essais cliniques de phase III des cancers bronchiques non à petites cellules avancés : une revue systématique**

### **Résumé**

#### **Objectifs**

L'objectif de cette étude est d'évaluer le report des critères de jugement de type temps à évènement (TAE) dans les essais cliniques de phase III des cancers bronchiques non à petites cellules avancés.

#### **Méthodes**

Une recherche dans la base de données MEDLINE a permis d'identifier les essais cliniques de phase III des cancers bronchiques non à petites cellules avancés publiés entre Janvier 2008 et Décembre 2012. Les articles éligibles étaient ceux qui reportaient des critères de jugement de type à évènement.

#### **Résultats**

Quarante et un articles ont été sélectionnés. 98 TAE ont été utilisés. La survie globale était le TAE le plus utilisée. Le critère de jugement principal était spécifié dans les 41 articles (100%) : 29 SG (73%), 7 survie sans-progression (17%), 1 survie sans toxicité de grade  $\frac{3}{4}$  (2%), 1 temps à échec du traitement (2%), 1 taux de réponse (2%), 1 qualité de vie (2%), 1 qualité de vie entre baseline et 18<sup>ème</sup> semaine (2%). Dans la partie « méthode », nous avons observé que la « date d'origine », « l'évènement » et « le mode de censure » étaient clairement définis dans 59 (60%), 50 (51%), et 11 (11%) des critères de jugement respectivement. Les TAE étaient clairement définis à la fois pour la « date d'origine » et « l'évènement d'intérêt » pour 43 critères de jugement (44%). Nous avons observé une tendance à l'augmentation du taux de critères de jugement bien définis au cours des années : 40%, 100%, 43%, 73%, 86% in 2008, 2009, 2010, 2011 et 2012 respectivement (p=0.058).

#### **Conclusion**

Une majorité des articles n'a pas fourni un report complet des TAE. Des recommandations sur le report des TAE dans les essais cliniques de phase III des cancers bronchiques non à petites cellules avancés sont nécessaires.

# **Time-to-event Endpoints Reporting in Phase III Advanced Non-Small-Cancer**

## **Clinical Trials: A Systematic Review**

Frédéric Fiteni (1,4), Guillaume Mouillet (1,4), Dewi Vernerey (1), Virginie W. Franck Bonnetain (1,2,3,5)

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

2 EA 3181 University of Franche-Comté, Besançon, France

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

4 University Hospital of Besançon, Department of Medical Oncology, Besançon

5 EORTC QOL Group, Brussels, Belgium

6 University Hospital of Besançon, Chest disease Department, Besançon, France

**Key words:** endpoint, Methodology, lung cancer

Corresponding Author:

Dr Frédéric Fiteni

2 place Saint Jacques 25 030 Besançon, France, Tel: + 33 (0)3 81 21 88 97

E-mail : [fredericfiteni@gmail.com](mailto:fredericfiteni@gmail.com)

Running head : Time-to-event endpoints in lung cancer trials



## **ABSTRACT**

### **Purpose**

The aim of the study was to evaluate the reporting of time-to event (TTE) endpoints in phase III advanced non-small cell lung cancer (NSCLC) clinical trials.

### **Methods**

A search in MEDLINE databases identified published phase III clinical trials in advanced NSCLC, published between January 2008 to December 2012. Eligible articles were those that reported time to event endpoints.

### **Results**

A total of 41 articles were selected. Among these trials, a total of 98 TTE endpoints were recorded. Overall survival (OS) was the most frequent TTE term, reported in all trials. The primary endpoint was specified in all the 41 studies (100%): 29 OS (73%), 7 progression-free survival (17%), 1 survival without grade 3/4 toxicity (2%), 1 time to treatment failure (2%), 1 overall response rate (2%), 1 quality of life (2%), 1 quality of life between baseline and the 18th week (2%). In the Methods section, we observed that the “starting point”, “events”, “information on censoring” were clearly defined for 59 (60%), 50 (51%), and 11 (11%) endpoints respectively. TTE terms were totally clearly defined in the methods section for “the starting point” and “events of interest” key points for 43 endpoints (44%). We observed a trend in an increase of the rate of defined primary endpoints over the years: 40%, 100%, 43%, 73%, 86% in 2008, 2009, 2010, 2011 and 2012 respectively (p=0.058).

### **Conclusion**

A majority of articles failed to provide a complete reporting of TTE. Guidelines for TTE reporting in phase III advanced NSCLC trials is warranted.

## **INTRODUCTION**

Cancer randomized clinical trials are conducted to obtain clinical evidence on the efficacy and safety of new interventions. The selection of appropriate endpoints is an important aspect of clinical trial design to achieve this objective (1). The publications of the CONSORT (Consolidated Standards of Reporting Trials) statement encouraged the reporting of clearly defined primary and secondary outcome measures specially time-to event (TTE) endpoints. This implies specifying the date of origin (time zero), the list of events to be considered as failures and the censoring process (2–4). Most of TTE are composite endpoints. Composite endpoints combine multiple events (called components) into a single endpoint. In oncology clinical trials, composite endpoints are time-to-event endpoints. An event is said to occur if any one of the prospectively defined components of the composite occurs. In the absence of at least one component considered in the composite endpoint, the patient is censored at the time of the last follow-up. The component must be of equal importance to the patient, similar frequency and similar reductions in relative risk (5). In a review of 125 randomized clinical trials in oncology, Mathoulin-Pelissier et al. showed that a majority of articles failed to provide a complete reporting of time-to-event (TTE) endpoints, thus introducing a source of uncontrolled variability (6). As an example, Birgisson et al. demonstrated that the inclusion of a second primary cancer other than the incident colorectal cancer as an event in the definition of DFS significantly impacted the results (7). The estimated DFS rate for patients with stage I-III disease was 62% after 5 years if this event was not counted as an event, compared with 58% if it was. The difference was larger for stage II (68 versus 60%) than for stage III disease (49 versus 47%). Another example is the PETACC 03 randomized study (8) where results were either significant or nonsignificant depending on whether second primary tumors were accounted for in the DFS definition. Nout et al. demonstrated that in- or exclusion of contralateral breast cancer or non-disease related death in the definition of recurrent disease or survival significantly affects estimated outcome probability. The estimation of 10-years DFS rate was 59.3% with contralateral breast cancer ignored and 55.5% with controlateral breast cancer as an event. The purpose of this study was to evaluate the reporting of time-to-event end points in phase III clinical trials of first-line chemotherapy in advanced NSCLC.

## **Methods**

## **Search Strategy and Selection for studies**

Eligible trials were randomized phase III trials of first-line chemotherapy in advanced NSCLC that reported primary analyses of TTE endpoints as primary or secondary objectives. Literature searches in PubMed database (January 2008 to December 2012) were performed. Searches were limited to English-language publications. For PubMed database research, the following strategies were used: (lung neoplasm[MeSH Terms]) AND (advanced[Text Word] OR metastatic[Text Word] + filters Clinical Trial, Phase III.

## **Data Extraction**

Two authors (F. F., G.M) independently extracted information using predefined data abstraction forms. Phase test sur x dossiers de la fiche de recueil. To assess interobserver reliability, one reviewer independently resolved disagreements (V.W and F.B). The following details were extracted: general items (number of patients, year of publication, study period, number of centers, nationality of the first author, academic or industrial trial); items related to each TTE endpoint and their section location in the articles (primary endpoint and its definition: date of origin, events, censoring process; secondary endpoints and their definitions; these items were based on CONSORT reporting items); other items related to methods of randomized clinical trials conduct (intent-to-treat-analysis [ITT], interim analysis, follow-up definition).

We assessed each TTE endpoint according to eight key parameters: definition of TTE clearly reported in the methods section (three key points: time of origin, events of interest, censoring events) and availability of TTE results (five key points: Kaplan-Meier estimation, precision (confidence interval), number of events, number of patients at risk, multivariate analysis). Each key point was coded as “yes,” “unclear,” or “no”.

## **Data Analyses**

We conducted a descriptive analysis of selected publications and of the TTE with each key point.

Quantitative variables were described with median and range. Qualitative variables were described with absolute frequencies (number) and relative frequencies (proportion).

To address the hypothesis that PFS definition in studies investigated is associated with distinct profiles, we performed an unsupervised hierarchical cluster analysis based on a combination of the starting point and events of interest for PFS definition.

The Cochran-Armitage trend test (9,10) was used to assess the evolution of the endpoints definition over the years ( $P \leq 0.05$  was considered significant).

Analyses were conducted with the use of SAS software, version 9.2 (SAS Institute) and R software (version 2.10.1). Hierarchical cluster analysis and dendrograms were performed with the cluster module of the amap package of R software (version 2.10.1).

## Results

Forty-one studies were included in the review (figure 1). Twenty-three trials were Europeans (56%), 11 Americans (27%) and 7 Asians (17%). There were 19 academic studies (46%) and 22 industrial studies (54%). The median number of including centers was 68 (range, 13 to 200 centers). The median number of patients was 561 (range, 165 to 1725).

Among the 41 articles, a total of 98 TTE endpoints were recorded and 39 were primary endpoints. Overall survival (OS) was the most frequent TTE term (41 terms, 42%). Other used terms were 31 progression-free survival (PFS) (32%), 10 time to progression (TTP) (10%), 6 duration of response (6%), 5 time to treatment failure (TTF) (5%), 1 duration of stable disease (1%), 1 time to deterioration of performance status (1%), 1 survival without grade 3/4 toxicity (1%), 1 survival without major grade 3/4 toxicity (1%), 1 survival without grade 4 toxicity (1%).

The primary endpoint was specified in all the 41 studies (100%): 29 OS (73%), 7 PFS (17%), 1 survival without grade 3/4 toxicity (2%), 1 TTF (2%), 1 overall response rate (2%), 1 quality of life (2%), 1 quality of life between baseline and the 18th week (2%) (overall response rate, quality of life and quality of life between baseline and the 18th week were not TTE endpoints).

In the Methods section, using the three key points to define TTE we observed that the “starting point”, “events”, “information on censoring” were clearly defined for 59 (60%), 50 (51%), and 11 (11%) endpoints respectively. TTE terms were totally clearly defined in the methods section for “the starting point” and “events of interest” key points for 43 endpoints (44%).

For OS, starting point was defined in 58% of cases (24/41) (54% randomization, 2% first chemotherapy, 2% date of registration) and event of interest defined in 61% of cases (25/41) (61% death from any cause). For PFS, progression was defined with RECIST criteria in 42% (13/31), with WHO criteria in 10% (3/31); clinical progression was used in one case and progression was not defined in 10% (13/31). For TTP, progression was defined with RECIST

criteria in 20% (2/10), WHO criteria in 20% (2/10) and progression was not defined in one case (table 1). Death was an event in one case. Radiological assessment was done by external independent radiologists in 6 trials (15%).

We observed a trend in an increase of the rate of defined primary endpoints over the years: 40%, 100%, 43%, 73%, 86% in 2008, 2009, 2010, 2011 and 2012 respectively ( $p=0.058$ ).

In the results section, using the five key points, the Kaplan-Meier estimation was clearly identified for 68 endpoints (69%), survival precision for 70 endpoints (71%), the number of patients at risk for 35 (36%), the number of events for 34 (35%), and the cox multivariate analysis for 25 (26%). We observed 6 endpoints (6%) with all key points clearly defined. The heat map (figure 2) shows two patterns of studies: studies without any definition of events of interest and starting point; and studies with definition of events of interest and starting point.

Follow-up was expressed in the results section for 29 TTE endpoints (30%) and for 5 (5%) of them the method of calculation was specified. A planned interim analysis was noted in 18 articles (44%), and the method used for the interim analysis was cited in 10 articles (24%).

## **Discussion**

To our knowledge, the present paper is the first to investigate TTE endpoint reported in phase III trials for advanced NSCLC. Our results show that a majority of phase III clinical trials for first-line chemotherapy in NSCLC used OS as primary endpoint (73%), followed by PFS (17%). The journals in which the trials have been published all emphasized the CONSORT statement (2–4), nevertheless 31 endpoints (32%) were not defined at all and only 43% of TTE were clearly defined. However a trend in improvement of the number of defined TTE ( $p=0.058$ ). Follow-up was expressed for only 30% of survival endpoints.

In 2008, these points were underlined by Mathoulin-Pelissier among 125 cancer randomized clinical trials (all cancers with 48% concerned in metastatic disease and 33% in the adjuvant setting) and particularly for the specification of the method for the assessment of events or the information on censoring. They highlighted that TTE were totally defined in the methods section for 113 end points (42%) which is similar to our results. Moreover they observed that the follow-up was expressed for only 57% of articles by the median follow-up without defining methods of calculation (6).

OS has been generally considered as the most relevant primary endpoint to address clinical benefit of treatment in oncology (1). Nevertheless, the increasing number of effective salvage treatments available in many types of cancer (ie subsequent lines of treatments) has resulted in the need for a larger number of patients to be included and/or the need of a more prolonged observation period to attain sufficient data that can achieve planned statistical power; this increases the cost of clinical trials and requires a longer duration to obtain results (1,11). Moreover, given the availability of other effective therapies administered after participation in the trial, interpretation of results can be confounded by the effects of post-trial therapy (12). Consequently endpoints such as PFS are often used because they are assessed earlier (ie intermediate endpoints) but these endpoints are tumor-centred endpoints (13,14) which might not directly reflect clinical benefit to the patient. Most of them are not systematically validated as surrogate endpoints for OS (15). In advanced NSCLC Laporte et al. conducted a meta-analysis from 5 trials and their analyses provided only modest support for considering PFS as an acceptable surrogate for OS in patients with advanced NSCLC (16). In our study, OS was the primary endpoint in 73% of trials and PFS in 17% of trials which underlined the necessity to improve the selection of an appropriate primary endpoint in advanced NSCLC phase III trials. For example, in EGFR mutated advanced NSCLC phase III trials, where long term survivors can be observed, PFS could be combined with Health-related quality of life as co-primary endpoints to ensure a clinical benefit of the treatment assessed.

Then, progression for PFS, TTP or TTF was mostly assessed by RECIST criteria but some trials used WHO criteria. RECIST and WHO criteria are two different methods to evaluate response in solid tumors. In RECIST criteria, measures are unidimensional and progression is defined as at least a 20% increase in the sum of diameters of target lesions, or new lesions (17). In WHO criteria, measures are bidimensional and progression is defined as at least a 25% increase in the sum of diameters of target lesions, or new lesions (18). Therefore, methods for assessment of events are a source of variability which limit cross comparison between studies.

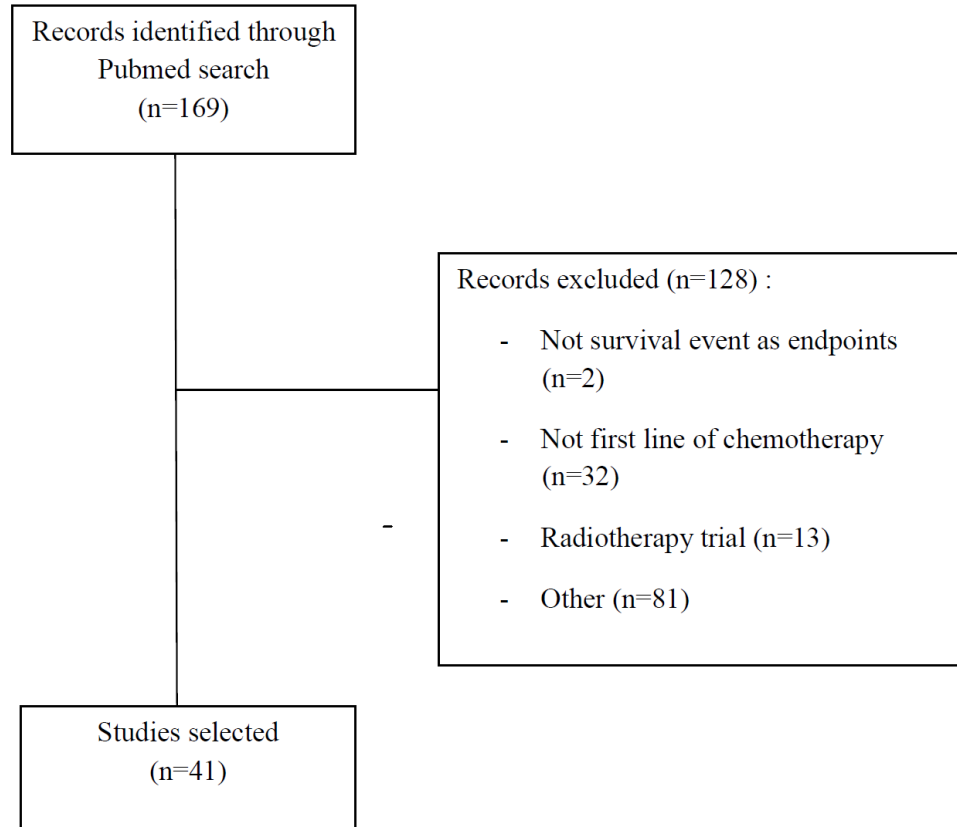
Another source of confusion can be the information on censoring which was defined for only 11% of endpoints. For example, patients alive without progression could be censored at the date of last news or at the date of the last clinical visit.

Furthermore we showed that PFS was defined in 86% of articles when PFS was the primary endpoint and in only 54% of cases when PFS was the secondary endpoint. Nevertheless the CONSORT statement request that primary and secondary outcomes must be clearly defined.

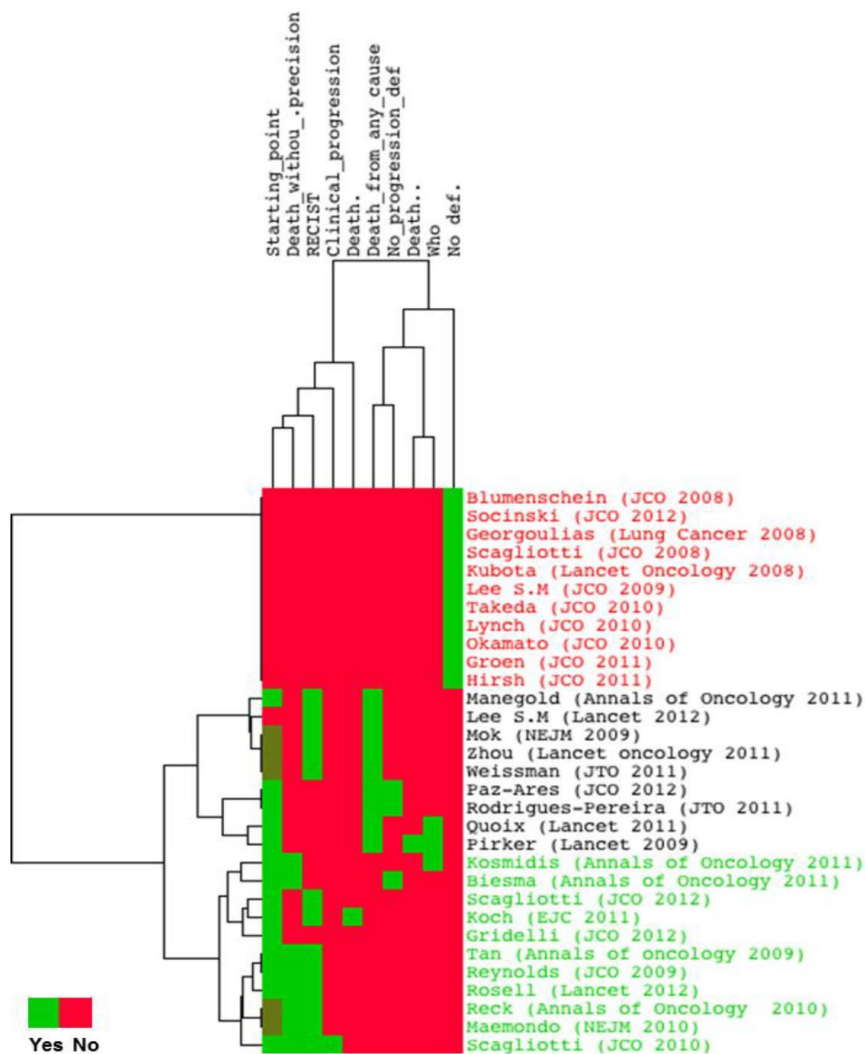
The limits of the articles length could explain the lack of definition of secondary endpoints. In this context, survival endpoints definitions could be provided in an international registration of clinical trials.

In the context of heterogeneity of definitions of survival endpoints, the international DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials) (19) was developed to obtain standardized consensus definitions for multiple cancer sites based on a formal consensus process and has already led to guidelines for time-to-event endpoint definitions in sarcomas and gastro-intestinal stromal tumors trials (20). Guidelines for survival endpoints reporting in phase III advanced NSCLC trials is warranted and is an objective of the DATECAN project.

**Figure 1.** Identification of phase III advanced non small cell lung cancer trials from PubMed search







**Figure 2** Distribution of Progression Free Survival (PFS) definition characteristics according to the 31 studies with PFS as an endpoint. The Heat map is colored according to positive (dark green) or negative (red) status for events of interest: RECIST criteria, WHO criteria, clinical progression, death from any cause, death from any cause when death occurred within 60 days of the last tumor response assessment or randomization (death.) and the starting point (no definition: red, first chemotherapy: light green and randomisation: dark green).

**Table 1.** Reported Definitions for the Time-to-Event End Points according to starting point and event of interest

Endpoint	Starting point	Event of interest
Overall survival (n=41)	Randomisation (n=22, 54%) First chemotherapy (n=1, 2%) Date of registration (n=1, 2%) No definition (n=17, 37%)	Death from any cause (n=10, 24%) Death without precision (n=15, 37%) No definition (n=16, 39%)
Progression free survival (n=31)	Randomisation (n=19, 58%) First chemotherapy (n=1, 3%) No definition (n=11, 39%)	RECIST progression (n=13, 42%) WHO progression (n=3, 10%) Clinical progression (n=1, 3%) No definition of progression (n=13, 10%) Death from any cause (n=9, 29%) Death without any documented progression (n=1, 3%) Death without precision (n=8, 26%) Death from any cause when death occurred within 60 days of the last tumour response assessment or randomisation (n=1, 3%) No definition (n=11, 35%)
Time to progression (n=10)	Randomisation (n=4, 40%) Date of registration (n=1, 10%) No definition (n=5, 50%)	RECIST progression (n=2, 20%) WHO progression (n=2, 20%) No definition of progression (n=1, 10%) Death without precision (n=1, 10%) No definition (n=5, 50%)
Duration of response (n=6)	First documented response (n=1, 17%) First complete or partial response (n=2, 33%) Time of the first tumor response (n=1, 17%) No definition (n=2, 33%)	RECIST progression (n=2, 33%) No definition of progression (n=2, 33%) Death from any cause (n=1, 17%) New cancer therapy (n=1, 17%) Death without precision (n=1, 17%) No definition (n=2, 33%)
Time to treatment failure (n=5)	Randomisation (n=3, 60%) No definition (n=2, 40%)	RECIST progression (n=2, 40%) Death from any cause (n=1, 20%) withdrawal due to adverse events (n=3, 60%) Failure to return (n=1, 20%) treatment refusal (n=2, 40%) New treatment started (n=3, 60%) Death without precision (n=2, 40%) No definition of progression (n=2, 60%) Loss to follow-up (n=1, 20%) WHO progression (n=1, 20%) Treatment discontinuation (n=1, 20%) No definition (n=1, 20%)
Duration of stable disease (n=1)	Randomisation	First documentation of disease progression in patients who did not achieve a best response of complete response or partial response
Time to deterioration of performance status (n=1)	No definition	ECOG performance status score of 3 or more
Survival without grade 3/4 toxicity (n=1)	Randomisation	The first grade 3 or 4 toxicity treatment-emergent adverse event that occurred within 30 days of the last treatment dose or death due to any cause whichever occurred first
Survival without important grade 3/4 toxicity (n=1)	Randomisation	Grade 3 or 4: neutropenia lasting more than 5 days, febrile neutropenia, neutropenia-related infections, anemia, thrombocytopenia, fatigue, nausea, vomiting, diarrhea, stomatitis, neurosensory events Death from any cause

## References

1. Fiteni F, Westeel V, Pivot X, Borg C, Vernerey D, Bonnetain F. Endpoints in cancer clinical trials. *J Visc Surg*. févr 2014;151(1):17-22.
2. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet*. 14 avr 2001;357(9263):1191-4.
3. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg Lond Engl*. 2012;10(1):28-55.
4. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med*. 17 avr 2001;134(8):663-94.
5. Montori VM, Permanyer-Miralda G, Ferreira-González I, Busse JW, Pacheco-Huergo V, Bryant D, et al. Validity of composite end points in clinical trials. *BMJ*. 12 mars 2005;330(7491):594-6.
6. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 août 2008;26(22):3721-6.
7. Birgisson H, Wallin U, Holmberg L, Glimelius B. Survival endpoints in colorectal cancer and the effect of second primary other cancer on disease free survival. *BMC Cancer*. 2011;11:438.
8. Van Cutsem E, Labianca R, Bodoky G, Barone C, Aranda E, Nordlinger B, et al. Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 juill 2009;27(19):3117-25.
9. Some Methods for Strengthening the Common  $\chi^2$  Tests on JSTOR [Internet]. [cité 24 févr 2015]. Disponible sur: <http://www.jstor.org/stable/3001616?origin=crossref>
10. Tests for Linear Trends in Proportions and Frequencies on JSTOR [Internet]. [cité 24 févr 2015]. Disponible sur: <http://www.jstor.org/stable/3001775?origin=crossref>
11. Saad ED, Buyse M. Overall survival: patient outcome, therapeutic objective, clinical trial end point, or public health measure? *J Clin Oncol Off J Am Soc Clin Oncol*. 20 mai 2012;30(15):1750-4.
12. Saad ED, Buyse M. Overall survival: patient outcome, therapeutic objective, clinical trial end point, or public health measure? *J Clin Oncol Off J Am Soc Clin Oncol*. 20 mai 2012;30(15):1750-4.

13. Ohorodnyk P, Eisenhauer EA, Booth CM. Clinical benefit in oncology trials: is this a patient-centred or tumour-centred end-point? *Eur J Cancer Oxf Engl* 1990. sept 2009;45(13):2249-52.
14. Booth CM, Ohorodnyk P, Eisenhauer EA. Call for clarity in the reporting of benefit associated with anticancer therapies. *J Clin Oncol Off J Am Soc Clin Oncol*. 20 nov 2009;27(33):e213-4.
15. Ciani O, Davis S, Tappenden P, Garside R, Stein K, Cantrell A, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. *Int J Technol Assess Health Care*. juill 2014;30(3):312-24.
16. Laporte S, Squifflet P, Baroux N, Fossella F, Georgoulas V, Pujol J-L, et al. Prediction of survival benefits from progression-free survival benefits in advanced non-small-cell lung cancer: evidence from a meta-analysis of 2334 patients from 5 randomised trials. *BMJ Open*. 2013;3(3).
17. Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer Oxf Engl* 1990. janv 2009;45(2):228-47.
18. Park JO, Lee SI, Song SY, Kim K, Kim WS, Jung CW, et al. Measuring response in solid tumors: comparison of RECIST and WHO response criteria. *Jpn J Clin Oncol*. oct 2003;33(10):533-7.
19. Bellera CA, Pulido M, Gourgou S, Collette L, Doussau A, Kramar A, et al. 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 Oxf Engl* 1990. mars 2013;49(4):769-81.
20. Bellera CA, Penel N, Ouali M, Bonvalot S, Casalli PG, Nielsen OS, et al. Guidelines for time-to-event endpoint definitions in sarcomas and gastro-intestinal stromal tumors (GIST) trials. Results of the DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann Oncol Off J Eur Soc Med Oncol ESMO*. 28 juill 2014;

## **Les critères de substitution de la survie globale dans les essais cliniques de cancer pulmonaire : une revue de littérature**

### **Résumé**

### **Objectifs**

L'objectif était de réaliser une revue de littérature des études qui ont évalué de potentiels critères de substitution de la SG dans les essais cliniques de cancer pulmonaire.

### **Methodes**

Une recherche des études ayant évalué de potentiels critères de substitution de la survie globale dans les essais cliniques de cancer pulmonaire a été réalisée sur la base de données PubMed. Nous avons évalué le caractère substitutif des potentiels critères de substitution par la « German institute of Quality and efficiency in Health Care's (IWQiG) framework » et la classification de Fleming.

### **Resultats**

Dix-huit études ont été identifiées. Pour les CBNPC opérables (essais adjuvants) et les CBNPC localement avancés (essais de radiothérapie), une seule méta-analyse sur données individuelles a retrouvé une forte corrélation entre la survie sans maladie et la survie sans progression avec la SG au niveau individuel et au niveau de l'essai clinique. Dans les essais de chimiothérapie adjuvante, le taux de corrélation entre la survie sans maladie et la SG était 0,83 au niveau individuel (IC 95%, 0,83–0,83) et 0,92 au niveau de l'essai clinique (IC 95% 0,88–0,95). Dans les essais de maladie localement avancée, le taux de corrélation entre la survie sans progression et la SG était de 0,77 à 0,85 (min-max) au niveau individuel, et 0,89 à 0,97 (min-max) au niveau de l'essai clinique. Cette étude fournit « une preuve » de la substitutivité de la survie sans maladie et la survie sans progression sur la SG selon la IWQiG framework et substitutivité était classée de niveau 2 dans la classification de Fleming.

### **Conclusion**

Le niveau de preuve montrant une relation entre SG et de potentiel critère de substitution dans les essais cliniques de cancer pulmonaire est faible sauf pour la survie sans maladie et la survie sans progression dans les CBNPC opérables (essais adjuvant) et localement avancés (essais radiothérapie). La SG reste le critère de jugement standard et la QdV constitue une mesure utile notamment lorsque la SG est difficile à mesurer.

## **SURROGATE ENDPOINTS FOR OVERALL SURVIVAL IN LUNG CANCER TRIALS: A REVIEW**

Frédéric Fiteni (1,2), Virginie Westeel (3), Franck Bonnetain (1,4,5)

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

2 University Hospital of Besançon, Department of Medical Oncology, Besançon, France

3 University Hospital of Besançon, Chest disease Department, Besançon, France

4 EA 3181 University of Franche-Comté, Besançon, France

5 The French National clinical research Platform Quality of Life and Cancer, Besançon France

Corresponding author: Frédéric Fiteni.

Address: University Hospital of Besançon, Department of Methodology and Quality of life in oncology unit, 3 boulevard Fleming, 25000, Besançon, France

E-mail address: [fredericfiteni@gmail.com](mailto:fredericfiteni@gmail.com)

Phone number: (33)3-81668796

Key words: lung cancer, surrogate endpoints

Short title: surrogate endpoints for overall survival

## **Abstract**

### **Background**

Our aim was to review the studies which assessed potential surrogate endpoints for overall survival (OS) in lung cancer trials.

### **Methods**

A Literature search in PubMed database of studies which assessed potential surrogate endpoints for OS in lung cancer trials was conducted. The surrogacy was assessed with the German institute of Quality and efficiency in Health Care's (IWQiG) framework and the Fleming hierarchy.

### **Results**

Eighteen studies were identified. In operable non-small cell lung cancer (NSCLC) (adjuvant trials) and locally advanced NSCLC (radiotherapy trials), one individual-patient data meta-analysis found a high correlation of disease-free survival (DFS) and progression-free survival (PFS) with OS at patient and trial level. In trials of adjuvant chemotherapy, correlation between disease-free survival DFS and OS were 0.83 at the individual level (95% CI 0.83–0.83) and 0.92 at trial level (95% CI 0.88–0.95). In studies of locally advanced disease, correlation between PFS and OS was 0.77 to 0.85 (range) at the individual level, and 0.89 to 0.97 (range) at trial level. This study provides a “proof” of the surrogacy of PFS and DFS on OS according to the IWQiG framework and the surrogacy of PFS and DFS on OS was classified level 2 according to Fleming hierarchy. In all the other setting, no endpoint was judged to be valid surrogate for OS.

### **Conclusion**

The level of evidence available supporting a relationship between OS and potential surrogate endpoints in lung cancer trials is low expect for DFS and PFS in operable NSCLC and locally advanced NSCLC.

## **Introduction**

As the American Society of Clinical Oncology stated, active treatment in cancer is generally undertaken with the goal of providing improved quantity and/or quality of patient survival (1). The selection of an appropriately valid primary endpoint is an important aspect of clinical trial design to achieve this objective. The Food and Drug Administration (FDA) considers OS benefit as the foundation for the approval of new anticancer drugs in the United States (2). Nevertheless, the increasing number of effective salvage treatments available in many types of cancer (ie subsequent lines of treatments) has resulted in the need for a larger number of patients to be included and/or the need of a more prolonged observation period to attain sufficient events that can achieve planned statistical power; this increases the cost of clinical trials and requires a longer duration to obtain results (3). Consequently intermediate endpoints such as progression-free survival are often used as primary endpoints because they are assessed earlier. However there is a lack of consistency in their definitions (4) and they are not systematically validated as surrogate endpoints for OS. Many methods have been developed to validate surrogate endpoints. Buyse and colleagues have proposed individual data meta-analysis and calculation of both the coefficient of determination ( $R^2$ ) individual and  $R^2$  trial to be the gold standard approach to the statistical surrogate validation. Based on the results of previously completed trials and using individual data, this approach jointly estimates: the correlation between the candidate surrogate (called ‘individual-level’ surrogacy) and the final endpoints and the correlation between the treatment effect on the candidate surrogate and its effect on OS (called ‘trial-level’ surrogacy).

Globally, in 2012, 1.8 million people were diagnosed with lung cancer, representing 12.9% of all cancers (5). Treatment for patients with lung cancer is dependent on the histological subtype and genetic subtype of the patient. Recently, improvement in the treatment of patients with advanced lung cancer, especially with the tyrosine kinase inhibitors has increased their survival. Therefore, lung trials may be conducted with intermediate endpoints as primary outcomes.

We conducted a review of studies which assessed potential surrogate endpoints for OS in lung cancer trials. We assessed the suitability of the potential surrogates using two validation framework: the Fleming criteria and the German Institute of Quality and efficiency in Health Care’s (IQWiG) framework.



## **Methods**

### **Search Strategy and Selection for studies**

A Literature searches in PubMed (from 1966 to May 2015) database were performed. Searches were limited to English-language publications. For PubMed database research, the following strategies were used: ("surrogate"[All Fields] AND "endpoints"[All Fields]) OR "surrogate endpoints"[All Fields] AND ("neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All Fields]). References of articles were also examined.

### **Data extraction**

The following details from the studies were extracted: authors, date of publication, number of patients included, type of lung cancer, stage of lung cancer, type of treatment, details of statistical methods to assess the association between the potential surrogate and the final endpoint, the results of these analyses, author's conclusions. The suitability of each endpoints as potential surrogate for OS in each study was analyze according to the Fleming criteria and the IQWiG framework. These criteria were applied by two reviewers and checked by a third reviewer.

### **German Institute of Quality and efficiency in Health Care framework**

The IQWiG an independent health technology assessment agency that assesses the benefits and harms of drug and non-drug technologies on behalf of the German Federal Joint Committee and the Federal Ministry of Health, published a framework for the validation of surrogate endpoints in oncology (6). The IQWiG framework proposes two levels of consideration to judge the suitability of a surrogate endpoint: the reliability of the evidence and the strength of evidence for surrogate validation, the strength-of-evidence criterion considers the degree of correlation of effects on the surrogate and the patient-relevant endpoint according to predefined thresholds (i.e., high correlation, when the lower limit of the 95 percent confidence interval for  $R \geq 0.85$ ; low correlation, when the upper limit of the 95 percent confidence interval for  $R \leq 0.7$ ; and medium correlation otherwise).

### **The Fleming criteria**

In 2005, Fleming proposed four level evidence hierarchy for potential surrogate endpoints: Level 1: a true clinical-efficacy measure; Level 2: a validated surrogate endpoint (for a specific disease setting and class of interventions); Level 3: a nonvalidated surrogate

endpoint, yet one established to be “reasonably likely to predict clinical benefit” (for a specific disease setting and class of interventions); and Level 4: a correlate that is a measure of biological activity but that has not been established to be at a higher level (7).

## **Results**

Eighteen studies were identified (figure 1). The table 1 summarizes the type of studies, the potential surrogate endpoints, the methodologies to assess surrogacy and the results. The assessment of the validity of potential surrogate endpoints according to IQWiG framework and the Fleming hierarchy as surrogate endpoint for OS in each study is presented in table 2.

### **Small-cell lung cancer**

Two studies, which assessed potential surrogate endpoints for OS in small-cell lung cancer trials, were identified.

In 2011, Foster et al. published an individual patient data (IPD) meta-analysis of 9 first-line extensive stage SCLC therapy trials (6 single-arm and 3 randomized trials). Patient-level associations showed that PFS status at 4 months was associated with OS (hazard ratio [HR], 0.42; 95% CI, 0.35-0.51; concordance index 0.63;  $p < 0.01$ ). At the trial level, correlation between PFS and OS was 0.79 (least squares  $R^2=0.79$ ) (8). This study provides a “hint” of an effect on OS given the effect observed on PFS. The surrogacy of PFS on OS was classified level 3 according to Fleming hierarchy.

In 2015, Imai et al retrospectively analyzed 49 cases of patients with SCLC who were treated with cisplatin and irinotecan as first-line chemotherapy (9). This study provides “no proof” of the surrogacy of an endpoint on OS according to the IQWiG framework.

### **Non-small cell lung cancer**

#### *Operable and locally advanced non-small cell lung cancer*

One study, which assessed surrogate endpoints for OS in trials for operable NSCLC (adjuvant trials) and locally advanced NSCLC (radiotherapy trials) was identified.

In 2013, Mauguen et al. conducted an IPD meta-analysis of 15 071 patients in 60 randomised trials. In trials of adjuvant chemotherapy, correlation between disease-free survival (DFS) and OS were 0.83 at the individual level (95% CI 0.83–0.83) and 0.92 at trial level (95% CI 0.88–0.95). In studies of locally advanced disease, correlation between PFS and OS was 0.77 to 0.85 (range, dependent on the regimen being assessed) at the individual level, and 0.89 to 0.97 (range) at trial level (10). This study provides a “proof” of the surrogacy of PFS and DFS on

OS according to the IWQiG framework and the surrogacy of PFS and DFS on OS was classified level 2 according to Fleming hierarchy.

*Advanced non-small cell lung cancer*

- Advanced Non-small cell lung cancer treated with epidermal growth factor receptor tyrosine kinase inhibitors

Three studies were identified: 2 retrospective studies (11,12) and 1 meta-analysis based on aggregate (13) (table 1). No endpoint was judged to be valid surrogate for OS according to IWQiG framework and the Fleming hierarchy (table 2).

- Advanced Non-small cell lung cancer treated with chemotherapy

Twelve studies were identified. Five of them assessed potential surrogate endpoints in first-line chemotherapy, one in maintenance chemotherapy two in second or third-line and four didn't specify the line of chemotherapy.

- First-line chemotherapy

Three meta-analyses based on aggregate data (14–16), one IPD meta-analysis (17) and one retrospective study (18) were conducted.

In 2013, Laporte et al. published an IPD meta-analysis of 2331 patients in 5 randomised trials. At patient-level associations, coefficient of determination between PFS and OS was 0.59 (95% CI, 0.58 to 0.61). At the trial level, the coefficient of determination between treatment effects on PFS and OS was 0.62 (95% CI, 0.52 to 0.72) (17). This study provides “no proof” of the surrogacy of PFS on OS according to the IWQiG framework and the surrogacy of PFS on OS was classified level 4 according to Fleming hierarchy.

In the three other studies, no endpoint was judged to be valid surrogate for OS according to IWQiG framework and the Fleming hierarchy.

- Maintenance chemotherapy

One study was identified. In 2013, Petrelli et al. conducted a meta-analysis on aggregate data. No endpoint was judged to be valid surrogate for OS according to IWQiG framework and the Fleming hierarchy (19).

- Second and third-line chemotherapy

Two studies were identified: one meta-analysis based on aggregate data (20) and one retrospective study (21). In these studies, no endpoint was judged to be valid surrogate for OS according to IWQiG framework and the Fleming hierarchy.

- Other studies

Three studies didn't specify the line of treatment: two meta-analysis based on aggregate data (22,23) and one IPD meta-analysis (24). (21). In these studies, no endpoint was judged to be valid surrogate for OS according to IWQiG framework and the Fleming hierarchy.

One meta-analysis on aggregate data pooled studies in first and second line of advanced NSCLC with targeted or standard chemotherapy. In this study, no endpoint was judged to be valid surrogate for OS according to IWQiG framework and the Fleming hierarchy

## **Discussion**

Our study is the first review specifically dedicated on surrogate endpoints in lung cancer trials. The statistical validation of surrogate endpoints has been a major focus of methodological researches over the last decade. A variety of methods have been proposed but at this time, no consensus exists with respect to the standards needed to identify valid surrogates (25,26). Nevertheless, Buyse and colleagues have developed a method based on individual data meta-analysis which assesses 'individual-level' surrogacy and 'trial-level' surrogacy which is considered as the gold standard (27). Currently, there is no comparison of methods based on IPD meta-analysis and on aggregate data meta-analysis. Linear regression between the true endpoint and the potential surrogate, based on small sample size retrospective study, is not recommended. A major difficulty for the validation of surrogate endpoints arises from the fact that they must be validated with respect to a specific therapeutic class, for a specific disease, at a specific stage. It is uncertain if the same surrogacy relationship will be applicable for a different treatment or at another stage of the disease. Then, since trials must be performed in order to validate a surrogate for a new therapeutic class, clinical benefit in trials using such endpoints remains uncertain (3,27).

In advanced SCLC, there was one IPD meta-analysis. Foster et al. demonstrated that PFS was associated with OS at the patient and trial levels (8). Nevertheless, a majority of single-arm trials were pooled in this meta-analysis. Moreover, individual level surrogacy was evaluated with a Cox proportional hazard ratio landmark analysis which is not the method developed by Buyse and colleagues. Furthermore, at the trial level, the  $R^2$  was 0.79 but the authors didn't provide the confidence interval as recommended by the IQWiG framework. Therefore, PFS could be a potential surrogate for OS in advanced SCLC but further validation is needed with data from a larger number of randomized phase III trials.

In operable NSCLC (adjuvant trials) and locally advanced NSCLC (radiotherapy trials), Mauguen et al. conducted an individual patient data meta-analysis and found a high

correlation of DFS and PFS with OS at patient and trial level (10). This result suggests that DFS and PFS in operable NSCLC (adjuvant trials) and locally advanced NSCLC (radiotherapy trials) could be considered as surrogate endpoint for OS. Nevertheless, extrapolation to other new therapies is not warranted.

In advanced NSCLC, there was only one IPD meta-analysis data assessed surrogacy of PFS at patient and trial levels (17). In this study the coefficients of determination were low.

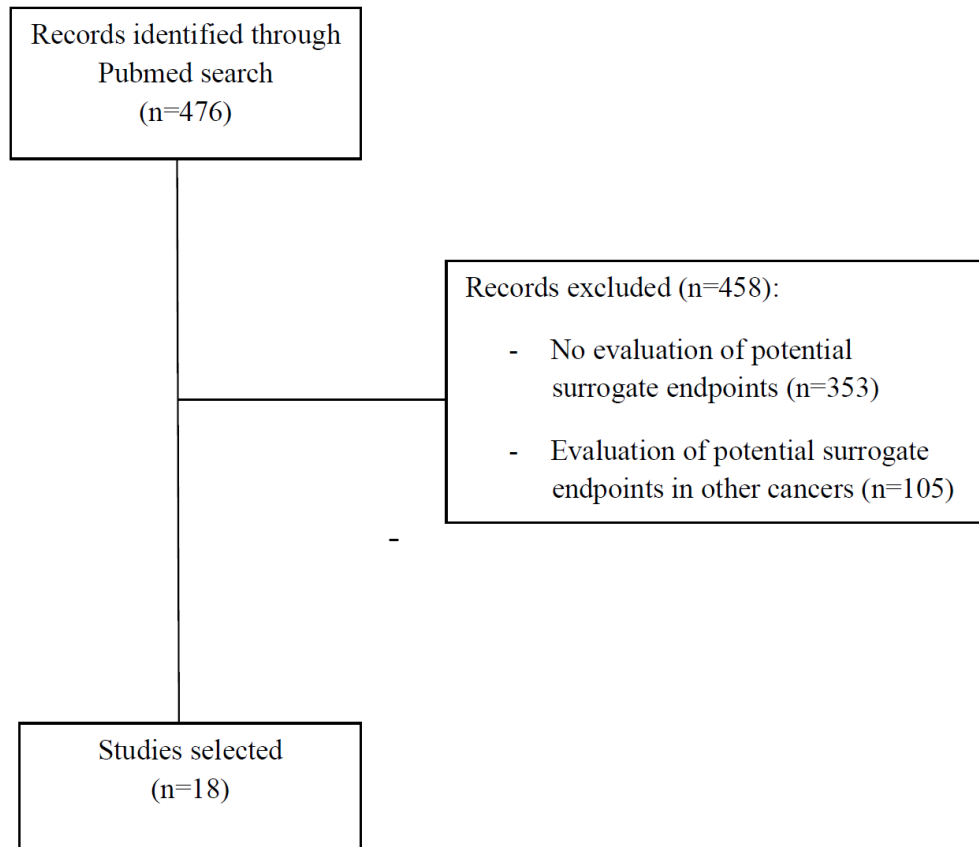
Therefore, no intermediate endpoint can be considered as a surrogate endpoint in patients with advanced NSCLC.

In summary, only one study provided evidence for the use of DFS and PFS in operable NSCLC (adjuvant trials) and locally advanced NSCLC (radiotherapy trials) as surrogate for OS, but confirmation of this result is necessary (10). There is no evidence in other setting. FDA guidance in NSCLC trials states that “Because of the subjectivity in the measurement of PFS assessments and the fact that assessments depend on frequency, accuracy, reproducibility, and completeness, the observed magnitude of effect should be substantial and robust” (28). The magnitude of effect is subjective and not clearly defined. Therefore, OS remains the most valid available endpoint in lung cancer trials.

Moreover, health-related quality of life (HRQoL) constitutes an alternative endpoint that allows earlier assessment of direct clinical benefit for the patient. Over the last decade, a tremendous growth in the incorporation of HRQoL in lung cancer clinical trials has been observed (3,29). The FDA considers HRQoL as an endpoint that assesses direct clinical benefit for the patient with NSCLC: “Patient-reported outcome (PRO) measures of tumor-related symptoms and functioning can provide direct evidence of treatment benefit if demonstrated to be well-defined and reliable assessments of a clinically meaningful concept or set of concepts, and if evaluated in well-conducted, placebo-controlled or double-blinded, randomized trials” (28). In assessing new therapies and approaches, recent authors have first published the results of efficacy and elsewhere the analysis of HRQoL: LUX-lung 3 (30,31), OPTIMAL (32,33), AVAPERL (34,35), LUX-lung 6 (36,37) trials . Another approach could be to combine intermediate endpoint such as PFS, with HRQoL as co-primary endpoints especially when treatments are associated with significant toxicity and surrogacy of other traditional efficacy clinical endpoints has not been demonstrated. Implementation of HRQoL as co-primary can provide a useful measure of the impact of intervention in a way that it is relevant to clinicians and patients especially when OS is difficult to measure (34).

In conclusion, we found that the level of evidence available supporting a relationship between OS and potential surrogate endpoints in lung cancer trials is low expect for DFS and PFS in operable NSCLC (adjuvant trials) and locally advanced NSCLC (radiotherapy trials). OS remains the gold standard and HRQoL can provide a useful measure of the impact of intervention in a way that it is relevant to clinicians and patients especially when OS is difficult to measure.

**Figure 1.** Identification of studies from PubMed search



**Table 1.** Surrogate endpoints for OS as reported in publications

Lung cancer setting	Author	Type of study	Methodology to assess surrogacy	Potential surrogates for OS (definitions)	Results
SCLC	Foster 2011	IPD meta-analysis from 870 patients participating in 6 single-arm (274 patients) and 3 randomized trials (596 patients)	ILS: Cox models using landmark analyses. TLS: R <sup>2</sup> from weighted least squares regression model, Spearman correlation coefficient, and R <sup>2</sup> from bivariate survival model (Copula R <sup>2</sup> ).	PFS	ILS: HR = 0.42; 95% CI, 0.35-0.51 TLS: R <sup>2</sup> from weighted least squares=0.79
				Best response Complete response Confirmed response	ILS: HR = 0.72; 95% CI, 0.61-0.85 TLS: R <sup>2</sup> from weighted least squares=0.21 ILS: HR = 0.65; 95% CI, 0.53-0.79 TLS: R <sup>2</sup> from weighted least squares=0.48 ILS: HR = 0.74; 95% CI, 0.62-0.88 TLS: R <sup>2</sup> from weighted least squares=0.50
	Imai 2015	Retrospective study of 49 patients	Spearman rank correlation analysis and linear regression analysis	PFS PPS Tumor shrinkage	r = 0.58, p < 0.05, R <sup>2</sup> = 0.24 r = 0.97, p < 0.05, R <sup>2</sup> = 0.94 r = 0.37, p < 0.05, R <sup>2</sup> = 0.13
Operable and locally advanced NSCLC	Mauguen 2013	IPD meta-analysis from 15 071 patients involved in 60 randomised clinical trials	ILS: R <sup>2</sup> from a bivariate survival model that takes censoring into account TLS: R <sup>2</sup> from a linear regression model, weighted by trial size	DFS PFS	ILS: R <sup>2</sup> =0.83, 95% CI 0.83–0.83 in trials without radiotherapy, and 0.87, 0.87–0.87 in trials with radiotherapy TLS: R <sup>2</sup> =0.92, 95% CI 0.88–0.95 in trials without radiotherapy and 0.99, 0.98–1.00 in trials with radiotherapy ILS: R <sup>2</sup> range 0.77–0.85, dependent on the regimen being assessed TLS: R <sup>2</sup> range 0.89–0.97
Advanced NSCLC treated with EGFR tyrosine kinase inhibitors	Tsujino 2009	Aggregate data meta-analysis from 24 phase II trials and 4 phase III trials with a total of 6171 patients	Correlation was examined using linear regression analysis.	Response rate Disease control rate	p < 0.0001 p < 0.003
	Li 2012	Aggregate data meta-analysis from 48 phase II trials and 12 phase III trials with a total of 9903 patients	Simple and multivariate linear regression models	Response rate Disease control rate PFS	R <sup>2</sup> = 0.71 (p < 0.0001) R <sup>2</sup> = 0.40 (p < 0.001) R <sup>2</sup> = 0.74 (p < 0.001)
	Yoshino 2014	Retrospective study of 35 patients	Spearman's rank correlation analysis and linear regression analysis	PFS PPS Tumor shrinkage	r=0.13, p=0.45, R <sup>2</sup> = 0.001 r=0.85, p < 0.05, R <sup>2</sup> =0.86 r=0.08, P=0.61, R <sup>2</sup> =0.01
	Hotta 2009	Aggregate data meta-analysis from 54 phase III trials with a total of 23457	Multiple linear regression analysis	Time to progression	R <sup>2</sup> =0.33
First line chemotherapy for advanced NSCLC	Hotta 2011	Aggregate data meta-analysis from 70 phase III trials with a total of 38721	Linear regression analysis, by assigning a weight equal to the sample size to each trial	SPP PFS	R <sup>2</sup> =0.8917 R <sup>2</sup> =0.2563
	Hayashi 2012	Aggregate data meta-analysis from 69 trials with a total of 37986 patients	Spearman's rank correlation coefficient	PPS PFS	r = 0.82, p < 0.0001 r = 0.43, p < 0.0001
	Laporte 2013	IPD meta-analysis from 2331 patients involved in 5 randomized clinical trials	ILS: bivariate copula model Kendall's τ was used to quantify the correlation between the endpoints. TLS: linear regression. Coefficients of determination (equal to squared correlation coefficients) were estimated using weighted linear regression.	PFS	ILS: τ=-0.59; 95% CI 0.58 to 0.61 TLS: R <sup>2</sup> =0.62, 95% CI 0.52 to 0.72
	Imai 2014	Retrospective study of 50 patients	Spearman rank correlation and linear regression analyses	PFS PPS Tumor shrinkage	r = 0.67, p < 0.05, R <sup>2</sup> = 0.39 r = 0.89, p < 0.05, R <sup>2</sup> = 0.79 r = 0.36, p < 0.05, R <sup>2</sup> = 0.14
	Petrelli 2013	Aggregate data meta-analysis from 10 phase III trials with a total of 4176 patients	Spearman rank correlation	PFS PPS	r= 0.37 r= 0.75
Second and third-line chemotherapy for advanced NSCLC	Hayashi 2013	Aggregate data meta-analysis from 18 phase III trials with a total of 11310 patients	Spearman's rank correlation coefficient	PFS PPS	r = 0.51 r = 0.94
	Imai 2014	Retrospective study of 39 patients	Spearman rank correlation and linear regression analyses	PFS PPS	r = 0.76, p < 0.05, R <sup>2</sup> = 0.50 r = 0.90, p < 0.05, R <sup>2</sup> = 0.85
Others studies for advanced NSCLC	Sekine 1999	Aggregate data meta-analysis from 102 phase II trials	Pearson's correlation coefficient	Response rate Rate of progressive disease	r = 0.62, p < 0.001 r = 0.80, p < 0.001
	Johnson 2006	Aggregate data meta-analysis from 191 randomised clinical trials	Linear regression weighted by trial	Response to treatment time to progression	R <sup>2</sup> = 0.16, p < 0.0001 R <sup>2</sup> = 0.19, p=0.0003
	Mandrekar 2010	IPD meta-analysis from 284 patients involved in 4 phase II trials	Univariable and multivariable Cox proportional hazards models, stratified for the patient's trial	PFS at 12 weeks FFS at 12 weeks	HR=0.39; 95% CI, 0.28-0.52 HR=0.37; 95% CI, 0.28-0.49
	Blumenthal 2015	Aggregate data meta-analysis from 14 phase III trials with a total of 12567 patients	weighted linear regression model	PFS ORR	R <sup>2</sup> = 0.89, 95% CI 0.80 to 0.98 R <sup>2</sup> =0.09, 95% CI, 0 to 0.33



**Abbreviations:** SCLC, small-cell lung cancer; NSCLC, non-small-cell lung cancer; EGFR, epidermal growth factor receptor; IPD, individual-patient data; ILS, individual-level surrogacy; TLS, trial-level surrogacy; PFS, progression-free survival; OS, overall survival; PPS, post-progression survival; DFS, disease-free survival; SPP, survival post progression; MST, median survival time; MPFS, median progression-free survival; FFS, failure-free survival; RR, response rate; PD, progression disease

**Table 2.** Assessment of the validity of intermediate endpoints as surrogate for OS

Lung cancer setting	Author	Endpoint	IQWiG framework			Fleming hierarchy
			Reliability	Correlation	Conclusion	
SCLC	Foster 2011	PFS	limited	medium	hint	level 3
		Best response	limited	low	no proof	level 4
		Complete response	limited	low	no proof	level 4
		Confirmed response	limited	low	no proof	level 4
	Imai 2015	PFS	low	NA	no proof	level 4
		PPS	low	NA	no proof	level 4
tumor shrinkage		low	NA	no proof	level 4	
operable and locally advanced NSCLS	Mauguen 2013	DFS	high	high	proof	level 2
		PFS	high	high	proof	level 2
Advanced NSCLC treated with EGFR tyrosine kinase inhibitors	Tsujiro 2009	response rate	moderate	NA	no proof	level 4
		disease control rate	moderate	NA	no proof	level 4
	Li 2012	response rate	moderate	medium	no proof	level 4
		disease control	moderate	low	no proof	level 4
	Yoshino 2014	PFS	moderate	medium	no proof	level 4
		PFS	low	low	no proof	level 4
		PPS	low	low	no proof	level 4
		tumor shrinkage	low	low	no proof	level 4
First line chemotherapy for advanced NSCLC	Hotta 2009	TTP	moderate	low	no proof	level 4
	Hotta 2011	SPP	moderate	low	no proof	level 4
		PFS	moderate	low	no proof	level 4
	Hayashi 2012	PPS	moderate	low	no proof	level 4
		PFS	moderate	low	no proof	level 4
	Laporte 2013	PFS	high	low	no proof	level 4
	Imai 2014	PFS	low	low	no proof	level 4
		PPS	low	low	no proof	level 4
Maintenance chemotherapy for advanced NSCLC	Petrelli 2013	PFS	moderate	low	no proof	level 4
		PPS	moderate	medium	no proof	level 4
Second and third-line chemotherapy for advanced NSCLC	Hayashi 2013	PFS	moderate	NA	no proof	level 4
		PPS	moderate	NA	no proof	level 4
	Imai 2014	PFS	low	NA	no proof	level 4
		PPS	low	NA	no proof	level 4
Others studies for advanced NSCLC	Sekine 1999	response rate	low	NA	no proof	level 4
		Rate of progressive disease	low	NA	no proof	level 4
	Johnson 2006	response to treatment	low	NA	no proof	level 4
		TTP	low	NA	no proof	level 4
	Mandrekar 2010	PFS at 12 weeks	moderate	NA	no proof	level 4
		FFS at 12 weeks	moderate	NA	no proof	level 4
Blumenthal 2015	PFS	moderate	low	no proof	level 4	
	ORR	moderate	low	no proof	level 4	

**Abbreviations:** SCLC, small-cell lung cancer; NSCLC, non-small-cell lung cancer; EGFR, epidermal growth factor receptor; PFS, progression-free survival; PPS, post-progression survival; DFS, disease-free survival; SPP, survival post progression; FFS, failure-free survival

## References

1. Peppercorn JM, Smith TJ, Helft PR, Debono DJ, Berry SR, Wollins DS, et al. American society of clinical oncology statement: toward individualized care for patients with advanced cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 20 févr 2011;29(6):755-60.
2. Beitz J, Gnecco C, Justice R. Quality-of-life end points in cancer clinical trials: the U.S. Food and Drug Administration perspective. *J Natl Cancer Inst Monogr*. 1996;(20):7-9.
3. Fiteni F, Westeel V, Pivot X, Borg C, Vernerey D, Bonnetain F. Endpoints in cancer clinical trials. *J Visc Surg*. févr 2014;151(1):17-22.
4. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 août 2008;26(22):3721-6.
5. Fact Sheets by Population [Internet]. [cité 9 juin 2015]. Disponible sur: [http://globocan.iarc.fr/Pages/fact\\_sheets\\_population.aspx](http://globocan.iarc.fr/Pages/fact_sheets_population.aspx)
6. A10-05 - Executive Summary 1.1 - Surrogate endpoints in oncology - A10-05\_Executive\_Summary\_v1-1\_Surrogate\_endpoints\_in\_oncology.pdf [Internet]. [cité 3 juin 2015]. Disponible sur: [https://www.iqwig.de/download/A10-05\\_Executive\\_Summary\\_v1-1\\_Surrogate\\_endpoints\\_in\\_oncology.pdf](https://www.iqwig.de/download/A10-05_Executive_Summary_v1-1_Surrogate_endpoints_in_oncology.pdf)
7. Fleming TR. Surrogate endpoints and FDA's accelerated approval process. *Health Aff Proj Hope*. févr 2005;24(1):67-78.
8. Foster NR, Qi Y, Shi Q, Krook JE, Kugler JW, Jett JR, et al. Tumor response and progression-free survival as potential surrogate endpoints for overall survival in extensive stage small-cell lung cancer: findings on the basis of North Central Cancer Treatment Group trials. *Cancer*. 15 mars 2011;117(6):1262-71.
9. Imai H, Mori K, Wakuda K, Ono A, Akamatsu H, Shukuya T, et al. Progression-free survival, post-progression survival, and tumor response as surrogate markers for overall survival in patients with extensive small cell lung cancer. *Ann Thorac Med*. mars 2015;10(1):61-6.
10. Mauguen A, Pignon J-P, Burdett S, Domerg C, Fisher D, Paulus R, et al. Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: a re-analysis of meta-analyses of individual patients' data. *Lancet Oncol*. juin 2013;14(7):619-26.
11. Tsujino K, Kawaguchi T, Kubo A, Aono N, Nakao K, Koh Y, et al. Response rate is associated with prolonged survival in patients with advanced non-small cell lung cancer treated with gefitinib or erlotinib. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. août 2009;4(8):994-1001.
12. Yoshino R, Imai H, Mori K, Takei K, Tomizawa M, Kaira K, et al. Surrogate endpoints for overall survival in advanced non-small-cell lung cancer patients with mutations of the epidermal growth factor receptor gene. *Mol Clin Oncol*. sept 2014;2(5):731-6.

13. Li X, Liu S, Gu H, Wang D. Surrogate end points for survival in the target treatment of advanced non-small-cell lung cancer with gefitinib or erlotinib. *J Cancer Res Clin Oncol*. nov 2012;138(11):1963-9.
14. Hotta K, Fujiwara Y, Matsuo K, Kiura K, Takigawa N, Tabata M, et al. Time to progression as a surrogate marker for overall survival in patients with advanced non-small cell lung cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. mars 2009;4(3):311-7.
15. Hotta K, Kiura K, Fujiwara Y, Takigawa N, Hisamoto A, Ichihara E, et al. Role of survival post-progression in phase III trials of systemic chemotherapy in advanced non-small-cell lung cancer: a systematic review. *PloS One*. 2011;6(11):e26646.
16. Hayashi H, Okamoto I, Morita S, Taguri M, Nakagawa K. Postprogression survival for first-line chemotherapy of patients with advanced non-small-cell lung cancer. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. juin 2012;23(6):1537-41.
17. Laporte S, Squifflet P, Baroux N, Fossella F, Georgoulas V, Pujol J-L, et al. Prediction of survival benefits from progression-free survival benefits in advanced non-small-cell lung cancer: evidence from a meta-analysis of 2334 patients from 5 randomised trials. *BMJ Open*. 2013;3(3).
18. Imai H, Takahashi T, Mori K, Ono A, Akamatsu H, Shukuya T, et al. Individual-level data on the relationships of progression-free survival, post-progression survival, and tumor response with overall survival in patients with advanced non-squamous non-small cell lung cancer. *Neoplasma*. 2014;61(2):233-40.
19. Petrelli F, Barni S. Is overall survival still the primary endpoint in maintenance non-small cell lung cancer studies? An analysis of phase III randomised trials. *Transl Lung Cancer Res*. févr 2013;2(1):6-13.
20. Hayashi H, Okamoto I, Taguri M, Morita S, Nakagawa K. Postprogression survival in patients with advanced non-small-cell lung cancer who receive second-line or third-line chemotherapy. *Clin Lung Cancer*. mai 2013;14(3):261-6.
21. Imai H, Mori K, Ono A, Akamatsu H, Taira T, Kenmotsu H, et al. Individual-level data on the relationships of progression-free survival and post-progression survival with overall survival in patients with advanced non-squamous non-small cell lung cancer patients who received second-line chemotherapy. *Med Oncol Northwood Lond Engl*. août 2014;31(8):88.
22. Sekine I, Tamura T, Kunitoh H, Kubota K, Shinkai T, Kamiya Y, et al. Progressive disease rate as a surrogate endpoint of phase II trials for non-small-cell lung cancer. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. juin 1999;10(6):731-3.
23. Johnson KR, Ringland C, Stokes BJ, Anthony DM, Freemantle N, Irs A, et al. Response rate or time to progression as predictors of survival in trials of metastatic colorectal cancer or non-small-cell lung cancer: a meta-analysis. *Lancet Oncol*. sept 2006;7(9):741-6.

24. Mandrekar SJ, Qi Y, Hillman SL, Allen Ziegler KL, Reuter NF, Rowland KM, et al. Endpoints in phase II trials for advanced non-small cell lung cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. janv 2010;5(1):3-9.
25. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med Res*. juin 2008;17(3):303-40.
26. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med*. 30 janv 2006;25(2):183-203.
27. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A. Biomarkers and surrogate end points--the challenge of statistical validation. *Nat Rev Clin Oncol*. juin 2010;7(6):309-17.
28. Clinical Trial Endpoints for the Approval of Non- Small Cell Lung Cancer Drugs and Biologics - UCM259421.pdf [Internet]. [cité 3 juin 2015]. Disponible sur: <http://www.fda.gov/downloads/Drugs/Guidances/UCM259421.pdf>
29. Fiteni F, Anota A, Westeel V, Bonnetain F. [Health-related quality of life in phase III cancer clinical trials: From questionnaire administration to statistical analysis]. *Bull Cancer (Paris)*. avr 2015;102(4):360-6.
30. Sequist LV, Yang JC-H, Yamamoto N, O'Byrne K, Hirsh V, Mok T, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol Off J Am Soc Clin Oncol*. 20 sept 2013;31(27):3327-34.
31. Yang JC-H, Hirsh V, Schuler M, Yamamoto N, O'Byrne KJ, Mok TSK, et al. Symptom control and quality of life in LUX-Lung 3: a phase III study of afatinib or cisplatin/pemetrexed in patients with advanced lung adenocarcinoma with EGFR mutations. *J Clin Oncol Off J Am Soc Clin Oncol*. 20 sept 2013;31(27):3342-50.
32. Zhou C, Wu Y-L, Chen G, Feng J, Liu X-Q, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol*. août 2011;12(8):735-42.
33. Chen G, Feng J, Zhou C, Wu Y-L, Liu X-Q, Wang C, et al. Quality of life (QoL) analyses from OPTIMAL (CTONG-0802), a phase III, randomised, open-label study of first-line erlotinib versus chemotherapy in patients with advanced EGFR mutation-positive non-small-cell lung cancer (NSCLC). *Ann Oncol Off J Eur Soc Med Oncol ESMO*. juin 2013;24(6):1615-22.
34. Barlesi F, Scherpereel A, Rittmeyer A, Pazzola A, Ferrer Tur N, Kim J-H, et al. Randomized phase III trial of maintenance bevacizumab with or without pemetrexed after first-line induction with bevacizumab, cisplatin, and pemetrexed in advanced

## **La qualité de vie relative à la santé dans les essais cliniques de phase III en oncologie : de l'administration du questionnaire à l'analyse statistique**

### **Résumé**

La QdV est un critère de jugement de plus en plus utilisé dans les essais de phase III en oncologie et qui est reconnu par la FDA ou l'ASCO comme un critère de jugement pour l'obtention des autorisations de mise sur le marché. Les cliniciens doivent donc prendre en compte les données de QdV dans leur prise de décision. Cet article expose les bases de l'étude de la QdV dans les essais de phase III en oncologie : les principaux questionnaires (FACT et EORTC-QLQ-C30) et les principales techniques d'analyse statistique : les modèles mixtes d'analyse de variance sur mesures répétées et le temps jusqu'à détérioration d'un score de QdV.



## La qualité de vie relative à la santé dans les essais cliniques de phase III en oncologie : de l'administration du questionnaire à l'analyse statistique

Frédéric Fiteni<sup>1,2,3</sup>, Amélie Anota<sup>1,3,4</sup>, Virginie Westeel<sup>5</sup>, Franck Bonnetain<sup>1,3,4,6</sup>

Reçu le 13 novembre 2014  
Accepté le 24 février 2015  
Disponible sur internet le :  
21 mars 2015

1. CHU de Besançon, unité de méthodologie en qualité de vie et cancérologie, 25000 Besançon, France
2. CHU de Besançon, service d'oncologie médicale, 25000 Besançon, France
3. Université de Franche-Comté, EA 3181, 25000 Besançon, France
4. Plateforme de qualité de vie et cancer, 25000 Besançon, France
5. CHU de Besançon, service de pneumologie, 25000 Besançon, France
6. EORTC QOL groupe, Bruxelles, Belgique

### Correspondance :

Frédéric Fiteni, CHU de Besançon, unité de méthodologie en qualité de vie et cancérologie, 25000 Besançon, France.  
[fredericfiteni@gmail.com](mailto:fredericfiteni@gmail.com)

### Mots clés

Qualité de vie relative à la santé  
Critère de jugement  
Méthodologie  
Essais cliniques  
Co-critère de jugement

### Résumé

Les critères de jugement font référence à des mesures cliniques ou biologiques afin d'évaluer l'efficacité de stratégies thérapeutiques. L'American Society of Clinical Oncology (ASCO) considère que le but d'un traitement actif en cancérologie est d'améliorer la quantité et/ou la qualité de vie des patients. La qualité de vie relative à la santé (QdV) reflète le jugement que porte le sujet sur sa propre santé qui comporte plusieurs dimensions à savoir physique, émotionnelle, sociale aussi bien que les symptômes dus à la maladie ou au traitement. La QdV est reconnue par l'ASCO et la FDA comme un critère de jugement pour l'autorisation de mise sur le marché des traitements en oncologie. De nombreux questionnaires d'autoévaluation de la QdV ont été développés et validés. Deux approches statistiques sont majoritairement utilisées pour l'analyse longitudinale. La première est l'utilisation des modèles linéaires mixtes sur mesures répétées. La seconde est une analyse de survie qui estime le temps jusqu'à détérioration de la QdV. Cependant, il n'existe aucun consensus sur les méthodes d'analyse et de report des modifications longitudinales des scores de QdV. Par ailleurs, la QdV pourrait être associée à d'autres critères de jugement comme la survie sans progression en tant que co-critères de jugement principaux mais l'utilisation des co-critères de jugement est une nouvelle approche dans les essais cliniques en oncologie qui nécessite des recherches méthodologiques pour promouvoir de tels schémas d'étude.

Keywords  
Health-related quality of life  
Endpoint  
Methodology  
Clinical trials  
Co-primary endpoint

## ■ Summary

### Health-related quality of life in phase III cancer clinical trials: From questionnaire administration to statistical analysis

*Endpoints refer to clinical and biological measurements that assess efficacy of therapeutic strategies. As the American Society of Clinical Oncology states, active treatment in cancer is generally undertaken with the goal of providing improved quantity and/or quality of patient survival. Health-related quality of life (HRQoL) reflects the patient-perceived evaluation of one's health, including physical, emotional, and social dimensions as well as symptoms due to disease or treatment. HRQoL is recognized as a component endpoint for cancer therapy approvals by the American Society of Clinical Oncology and the FDA. Many self-completion HRQoL questionnaires have been developed and validated. Two main statistical methods have been developed to longitudinally analyze HRQoL. The first is the linear mixed model for repeated measure (LMM). The second is a survival approach, which estimates the time to HRQoL deterioration. However, there is no guideline for methods of analyzing and reporting longitudinal changes in HRQoL scores. Moreover, HRQoL could also be combined with other endpoints like progression-free survival as co-primary endpoint, but the use of co-primary endpoints in cancer clinical trials is a new approach and methodological researches must be pursued to promote such designs.*

## Introduction

La qualité de vie relative à la santé (QdV) est reconnue par l'American Society of Clinical Oncology (ASCO) et la Food and Drug Administration (FDA) comme un critère de jugement des essais cliniques en oncologie [1]. Il existe un large consensus pour affirmer que la QdV est un concept multidimensionnel incluant au minimum les domaines physiques, psychiques et sociaux ainsi que les symptômes liés à la maladie et aux traitements. Cette définition découle indirectement de la définition de la santé de l'OMS [2].

En cancérologie, le critère de jugement de référence pour évaluer un nouveau traitement dans un essai de phase III est la survie globale (SG). Cependant, avec l'augmentation des thérapies efficaces contre beaucoup de cancers, un suivi plus long et un nombre plus important d'inclusions sont nécessaires pour observer un nombre de décès suffisants et atteindre la puissance statistique requise [3,4]. Par conséquent, de nombreux essais cliniques utilisent des critères de jugement intermédiaires évaluables plus précocement et qui requièrent moins de patients comme la survie sans progression (SSP). Néanmoins, ceux-ci n'ont pas été validés comme critères de substitution de la SG et présentent une hétérogénéité de définition entre les essais cliniques [5].

Dans ce contexte, les résultats de QdV dans les essais de phase III peuvent aider les instances à la prise de décision d'autorisation de mise sur le marché de nouveau traitement et aider les cliniciens à la prise de décision thérapeutique. Cependant, du fait de son caractère subjectif et multidimensionnel, son évaluation, analyse et interprétation des résultats demeurent complexes. De nombreux instruments de mesure de la QdV ont été développés et deux principales méthodes d'analyse

longitudinale sont utilisées : les modèles linéaires mixtes sur mesures répétées et le temps jusqu'à détérioration d'un score de QdV (TJD).

## Les questionnaires de qualité de vie relative à la santé en oncologie

Dans les essais de phase III en oncologie, les deux principaux questionnaires utilisés sont l'European Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire QLQ-C30 et le Functional Assessment of Cancer Therapy-General (FACT-G) [6,7]. Ces deux questionnaires sont spécifiques de l'oncologie. Ce sont des auto-questionnaires qui ont fait l'objet d'une validation selon des étapes spécifiques : validité de contenu, fiabilité (reproductibilité, consistance interne), sensibilité au changement, validation transculturelle [8].

Le QLQ-C30 est composé de 30 items permettant d'évaluer 15 dimensions de QdV : une dimension de santé globale, 5 dimensions de fonctionnement (physique, limitation des tâches habituelles, émotionnel, cognitif, social), 9 dimensions de symptômes (fatigue, nausée et vomissement, douleur, dyspnée, insomnie, perte d'appétit, constipation, diarrhée, difficulté financière). Chaque dimension (ou échelle) de QdV est mesurée par un ou plusieurs items (de 1 à 5 items). Par exemple, la dimension « perte d'appétit » est évaluée à partir de la réponse à l'item n° 13, la dimension « fatigue » à partir des items n°s 10, 12 et 18. Un score compris entre 0 et 100 est calculé pour chaque dimension à partir des réponses aux items. Pour les dimensions de fonctionnement ou de santé globale, un score élevé représente un bon fonctionnement ou une bonne santé globale et traduit donc une meilleure QdV, inversement pour les dimensions de symptômes un score élevé représente



une importante symptomatologie, ce qui traduit plutôt une moins bonne QdV. Des modules complémentaires spécifiques à différents cancers ont été développés, comme par exemple, le BR23 pour les cancers du sein ou le LC13 pour les cancers broncho-pulmonaires.

Le FACT-G est composé de 27 items à 5 modalités de réponse. Les questions sont subdivisées en 4 dimensions de QdV : le bien-être physique (7 items), le bien-être social/familial (7 items), le bien-être émotionnel (6 items) et le bien-être fonctionnel (7 items). Un score est calculé par dimension tel qu'un score élevé représente un niveau de bien-être élevé. Un score global est également calculé, correspondant à la moyenne des sous-échelles et reflétant un niveau de QdV global.

Les différences de conceptualisation des questionnaires entre les groupes EORTC et FACT sont résumées dans le [tableau 1](#).

À l'heure actuelle, il n'existe pas de recommandation pour la sélection du meilleur questionnaire selon la pathologie et le traitement étudié, ce qui peut rendre difficile la comparaison entre certains essais. Par exemple, les essais RTOG 0825 [9] et AVAglio [10] qui ont étudié l'intérêt d'associer le bévacicumab à la radiothérapie et au témozolomide dans le traitement de première ligne du glioblastome ont utilisé le questionnaire EORTC QLQ-C30 et le module complémentaire BN20 spécifiques des tumeurs cérébrales. A contrario, les essais OPTIMAL [11,12] et LUX-Lung 3 [13,14] qui ont comparé un inhibiteur à tyrosine kinase de l'EGFR (erlotinib pour OPTIMAL et afatinib pour

LUX-Lung 3) à une chimiothérapie en première ligne des cancers broncho-pulmonaires avancés EGFR mutés ont utilisé respectivement les questionnaires FACT-L (OPTIMAL) et EORTC QLQ-C30 (LUX-Lung 3).

Par conséquent, des recommandations devraient être établies pour la sélection des questionnaires, bien qu'il soit toujours nécessaire de développer de nouveaux questionnaires adaptés aux nouvelles thérapies et à leur impact sur la QdV.

De plus, il est essentiel de spécifier dans le protocole le rythme d'administration des questionnaires. L'EORTC recommande un minimum de 3 recueils : avant, pendant et après le traitement. Cependant, une évaluation plus fréquente est souvent nécessaire pour mettre en évidence une différence de QdV significative entre les bras de traitement. En outre, le rythme d'administration dépend du schéma de l'étude, des objectifs et des modalités de traitement.

Par ailleurs, une personne responsable de l'administration des questionnaires doit être clairement identifiée dans chaque centre. La plupart du temps il s'agit d'un attaché de recherche clinique. Cette personne doit savoir quand le patient arrive, s'assurer que le patient reçoive bien le questionnaire dans un endroit calme avant l'annonce d'un diagnostic ou une procédure thérapeutique et implémente les données de suivi au cas où le patient ne soit pas disponible. Il est important que l'administration des questionnaires se fasse les jours de traitement ou de consultation de suivi afin de limiter les données manquantes. La

**TABLEAU 1**  
**Différence de conceptualisation des questionnaires de qualité de vie relative à la santé entre les groupes EORTC et FACT**

	EORTC	FACT
Domaines	Centrés sur les symptômes et dimensions fonctionnelles	Dimensions 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

méthode d'administration la plus fréquemment utilisée est sur papier. L'aide de l'attaché de recherche clinique peut être utile si le patient présente des difficultés de lecture ou d'écriture. La procédure d'administration des questionnaires doit être décrite dans le protocole.

### Analyse statistique longitudinale de la qualité de vie relative à la santé

La première étape avant l'analyse longitudinale est la détermination des dimensions de QdV cibles et la différence minimale cliniquement importante (DMCI).

Tout ceci doit être établi lors de la conception de l'étude et clairement spécifié dans le protocole. La DMCI représente le plus petit changement de score qui est cliniquement pertinent. Une DMCI de 5 ou 10 points est couramment utilisée pour les dimensions du questionnaire EORTC QLQ-C30 [15,16] et 7 points pour le FACT-G [17,18]. Il n'existe aucune recommandation sur le choix de la DMCI. Deux principales techniques d'analyse longitudinale sont utilisées : les modèles mixtes d'analyse de variance sur mesures répétées et le TJD [19].

### Les modèles mixtes d'analyse de variance sur mesures répétées

Les modèles mixtes d'analyse de variance sur mesures répétées sont des modèles complexes qui permettent de modéliser des données telles que, pour chaque individu considéré, on dispose d'observations à différents instants, autrement dit répétées dans le temps. Dans le cadre d'une analyse longitudinale de QdV, ces modèles permettent d'estimer l'effet du temps (c'est-à-dire les changements de score au cours du temps), l'effet groupe (i.e. des différences de scores entre les bras de traitement à l'inclusion) et une interaction entre le bras de traitement et le temps (reflétant une évolution différente du score de QdV selon le bras de traitement au cours du temps comparativement au bras de référence). Ils contiennent des effets fixes comme le traitement et des effets aléatoires comme l'effet patient en prenant en compte la corrélation entre les mesures faites pour un même patient à différents moments. Pour cela, le type de matrice de covariance entre les mesures de QdV doit être précisé. Les structures de matrice les plus utilisées sont les non-structurées, auto-régressives d'ordre 1 et à symétrie composée. Les critères d'information Akaike Information Criteria (AIC) et Bayesian Information Criteria (BIC) sont utilisés pour sélectionner la meilleure structure.

Dans l'essai AURELIA [20], l'ajout du bévacizumab à une monochimiothérapie dans les cancers ovariens platine-résistants a montré une augmentation significative de la SSP (critère de jugement principal) (*hazard ratio* : 0,48 ; intervalle de confiance à 95 % : 0,38–0,60 ;  $p < 0,001$  ; médiane : 6,7 mois bras bévacizumab vs 3,4 mois bras chimiothérapie seule). Il n'y avait pas de différence entre les deux bras en termes de SG (*hazard ratio* : 0,85 ; intervalle de confiance à 95 % : 0,85–1,08 ;

$p = 0,174$  ; médiane : 16,6 mois bras bévacizumab vs 13,3 mois bras chimiothérapie seule). L'étude de l'impact sur la QdV était donc essentielle afin d'étudier le bénéfice clinique pour le patient et a été publié dans un second temps [21]. Les questionnaires utilisés étaient l'EORTC QLQ-C30 et son module complémentaire spécifique du cancer de l'ovaire OV28, ainsi que le FACT-Ovarian Cancer Symptom Index (FOSI). La DMCI était déterminée à 15 %, c'est-à-dire 15 points sur 100 pour les dimensions des questionnaires EORTC QLQ-C30 et OV28 et 5 points sur 32 pour le FOSI. La dimension « symptômes abdominaux/gastro-intestinaux » de l'OV28 était définie comme la dimension principale d'analyse. Plus de patients ont expérimenté une amélioration  $\geq 15$  % des symptômes abdominaux/gastro-intestinaux dans le bras bévacizumab à la semaine 8 (21,9 % vs 9,3 % ; différence : 12,7 % ; intervalle de confiance à 95 % : 4,4–20,9 ;  $p \leq 0,002$ ). Ces résultats viennent renforcer ceux obtenus sur la SSP alors que l'analyse en SG n'était pas significative.

L'approche par modèles linéaires mixtes d'analyse de variance sur mesures répétées présente l'avantage d'être bien établie mais nécessite que les scores aient une distribution normale, ce qui est rarement vérifié. Par ailleurs, ils ne prennent pas en compte le phénomène de changement de réponse (*response shift*) (changement de référentiel de perception de la QdV au cours du temps chez le patient) et sont difficiles d'interprétation pour les cliniciens (tableau II).

### Le temps jusqu'à détérioration d'un score de qualité de vie relative à la santé

L'analyse du TJD d'un score de QdV est une approche de type analyse de survie. Il peut être défini comme l'intervalle de temps entre la date de randomisation et l'observation d'une première diminution d'une valeur seuil de la DMCI d'un score de QdV par rapport au score à l'inclusion. De multiples définitions peuvent être proposées en fonction : de la prise en compte du décès comme évènement, du choix du score de référence (score à l'inclusion, score précédent, meilleur score précédent), de la valeur seuil de la DMCI [19]. L'estimation du TJD se fait généralement selon la méthode de Kaplan-Meier. Une comparaison entre les deux bras de traitement peut être réalisée selon le test du log-rank et un modèle de Cox multivarié peut être réalisé afin d'estimer la taille de l'effet traitement (*hazard ratio*).

L'essai BOLERO-2 [22] a démontré que l'ajout de l'évérolimus à l'exémestane chez les patientes ménopausées atteintes d'un cancer du sein avancé après progression sous anti-aromatases augmente de manière significative la SSP (critère de jugement principal) (11,0 mois versus 4,1 mois ; *hazard ratio* : 38 ; intervalle de confiance à 95 % : 0,31–0,48 ;  $p < 0,0001$ ). Cependant, davantage de toxicités de grade 3 et 4 pour le bras contenant l'évérolimus ont été observées (mucite, anémie, dyspnée, hyperglycémie, fatigue, pneumopathie). L'étude de l'impact du traitement sur la QdV était donc essentielle et a été publiée

TABLEAU II  
Avantages et inconvénients des techniques d'analyse longitudinale des données de qualité de vie relative à la santé

	Avantages	Inconvénients
Modèle mixte d'analyse de variance sur mesures répétées	Technique bien établie	Phénomène de changement de réponse non pris en compte
	Facilité de réalisation	Difficile à interpréter pour les cliniciens
	Robuste malgré les données manquantes	
Temps jusqu'à détérioration d'un score de qualité de vie relative à la santé	Résultats facilement interprétables	
	Possibilité de prendre en compte les données manquantes	Multiplicité des définitions
	Possibilité de prendre en compte le décès	Puissance statistique
	Possibilité de prendre en compte le phénomène de changement de réponse	

dans un second temps [23]. Le questionnaire utilisé était l'EORTC QLQ-C30. La DMCI était déterminée à 5 %. La dimension de « santé globale » était définie comme la dimension principale d'analyse. Les temps médians jusqu'à détérioration du score de la dimension « santé globale » étaient de 8,3 mois pour le bras avec évérolimus et 5,8 mois pour le bras sans évérolimus (*hazard ratio* : 0,74 ;  $p = 0,0084$ ).

Cette technique présente l'avantage d'être facile d'interprétation puisqu'elle utilise les méthodes classiques d'analyse de survie dont les cliniciens sont familiers, avec des résultats présentés de manière visuelle (courbe de Kaplan-Meier). Cependant, cette technique aurait une puissance statistique inférieure à celles des modèles mixtes d'analyse de variance [24] (tableau II). Enfin, il existe une multiplicité de définitions possibles du TJD et des recommandations semblables aux critères radiologiques RECIST devraient donc être développées [19].

### Design d'essais incluant la QdV comme critère de jugement

La QdV est la plupart du temps un critère de jugement secondaire dans les essais de phase III. De plus en plus d'auteurs publient dans un second temps l'analyse de QdV, par exemple, pour les cancers du sein : BOLERO-2 [22,23], EMILIA [25,26], CLEOPATRA [27,28] ; les cancers broncho-pulmonaires : LUX-lung 3 [13,14], OPTIMAL [11,12], AVAPERL [29,30] ; les cancers ovariens : AURELIA [20,21], GOG 0218 [31,32] ; les cancers pancréatiques : PRODIGE 4/ACCORD 11 [33,34]. Dans ce contexte, il est nécessaire de bien préciser dans le protocole les dimensions cibles de QdV et les méthodes d'ajustement du risque alpha, car il est bien connu en statistique que la multiplicité des tests augmente le risque d'erreur de type I [35].

Cependant, l'augmentation des publications des résultats de QdV souligne l'intérêt des auteurs à améliorer la pertinence des résultats des critères de jugement de type SSP. Néanmoins, ces essais n'étaient pas conçus méthodologiquement pour l'analyse de la QdV. Par conséquent, une approche innovante serait de combiner les critères de jugement intermédiaires type SSP avec la QdV en tant que co-critères principaux des études [3]. Dans ce cas, les procédures de contrôle du risque d'erreur de type I devrait être précisées dans le protocole et le calcul du nombre de sujet nécessaire devrait être adapté à l'augmentation du risque de type I. Des recherches méthodologiques doivent être poursuivies afin d'améliorer le design de ce type d'essais avec co-critères de jugement principaux.

### Conclusion

L'étude de la QdV dans les essais cliniques de phase III en oncologie est un critère de jugement de plus en plus important qui reflète un bénéfice clinique direct pour le patient. La QdV est reconnue par les instances comme un critère de jugement pour l'autorisation de mise sur le marché des traitements en oncologie. Néanmoins, les méthodes de mesure et d'analyse de la QdV sont complexes, hétérogènes et non standardisées. Les paramètres nécessaires à décrire dans la partie méthodologie des essais de phase III concernant l'analyse statistique de la QdV sont les suivants : les dimensions de QdV cibles, les hypothèses a priori de résultats de QdV, les méthodes d'ajustement du risque alpha, la DMCI, la population d'étude (*intention-to-treat*, *modified intention-to-treat*, etc.), la méthode de prise en compte des données manquantes, la méthode d'analyse longitudinale (TJD et/ou modèles mixtes d'analyse de variance sur mesures répétées) et sa définition. De plus, l'utilisation à la fois de la technique du TJD et des modèles mixtes pour l'analyse des

données de QdV d'un essai permettrait d'évaluer la sensibilité des résultats et d'apporter des éléments de discussion sur l'interprétation des résultats. En effet, la technique des modèles mixtes d'analyse de variance sur mesures répétées est une technique bien établie, facile à mettre en œuvre mais est difficile d'interprétation et ne prend pas en compte le phénomène de changement de réponse. La technique du TJD présente des résultats faciles à appréhender et peut prendre en compte le phénomène de changement de réponse, mais il existe une multiplicité de

définitions possibles du TJD et semblerait avoir une puissance statistique moins importante que la technique des modèles mixtes d'analyse de variance sur mesures répétées.

L'ensemble de ces éléments souligne la nécessité de poursuivre les recherches méthodologiques afin de standardiser l'analyse de QdV et permettre une meilleure analyse conjointe des essais.

**Déclaration d'intérêts :** les auteurs déclarent ne pas avoir de conflits d'intérêts en relation avec cet article.

## Références

- [1] Peppercorn JM, Smith TJ, Helft PR, Debono DJ, Berry SR, Wollins DS, et al. American society of clinical oncology statement: toward individualized care for patients with advanced cancer. *J Clin Oncol* 2011;29(6):755-60.
- [2] World Health Organization. WHO Constitution. Geneva: WHO; 1948.
- [3] Fiteni F, Westeel V, Pivot X, Borg C, Vernerey D, Bonnetain F. Endpoints in cancer clinical trials. *J Visc Surg* 2014;151(1):17-22.
- [4] Saad ED, Buyse M. Overall survival: patient outcome, therapeutic objective, clinical trial end point, or public health measure? *J Clin Oncol* 2012;30(15):1750-4.
- [5] Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J Clin Oncol* 2008;26(22):3721-6.
- [6] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, 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):365-76.
- [7] Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol* 1993;11(3):570-9.
- [8] Sprangers MA, Groenvold M, Arraras JL, Franklin J, te Velde A, Muller M, et al. 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* 1996;14(10):2756-68.
- [9] Gilbert MR, Dignam JJ, Armstrong TS, Wefel JS, Blumenthal DT, Vogelbaum MA, et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N Engl J Med* 2014;370(8):699-708.
- [10] Chinot OL, Wick W, Mason W, Henriksson R, Saran F, Nishikawa R, et al. Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N Engl J Med* 2014;370(8):709-22.
- [11] Chen G, Feng J, Zhou C, Wu Y-L, Liu X-Q, Wang C, et al. Quality of life (QoL) analyses from OPTIMAL (CTONG-0802), a phase III, randomised, open-label study of first-line erlotinib versus chemotherapy in patients with advanced EGFR mutation-positive non-small-cell lung cancer (NSCLC). *Ann Oncol* 2013;24(6):1615-22.
- [12] Zhou C, Wu Y-L, Chen G, Feng J, Liu X-Q, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* 2011;12(8):735-42.
- [13] Wu Y-L, Zhou C, Hu C-P, Feng J, Lu S, Huang Y, et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol* 2014;15(2):213-22.
- [14] Yang J-C-H, Hirsh V, Schuler M, Yamamoto N, O'Byrne KJ, Mok TSK, et al. Symptom control and quality of life in LUX-Lung 3: a phase III study of afatinib or cisplatin/pemetrexed in patients with advanced lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013;31(27):3342-50.
- [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* 1998;16(1):139-44.
- [16] Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *J Clin Oncol* 2011;29(1):89-96.
- [17] Eton DT, Fairclough DL, Cella D, Yount SE, Bonomi P, Johnson DH, et al. Early change in patient-reported health during lung cancer chemotherapy predicts clinical outcomes beyond those predicted by baseline report: results from Eastern Cooperative Oncology Group Study 5592. *J Clin Oncol* 2003;21(8):1536-43.
- [18] Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C, et al. What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. *J Clin Epidemiol* 2002;55(3):285-95.
- [19] Anota A, Hamidou Z, Paget-Bailly S, Chibaudel B, Bascoul-Mollevis C, Auquier P, 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.
- [20] Pujade-Lauraine E, Hilpert F, Weber B, Reuss A, Poveda A, Kristensen G, et al. Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: the AURELIA open-label randomized phase III trial. *J Clin Oncol* 2014;32(13):1302-8.
- [21] Stockler MR, Hilpert F, Friedlander M, King MT, Wenzel L, Lee CK, et al. Patient-reported outcome results from the open-label phase III AURELIA trial evaluating bevacizumab-containing therapy for platinum-resistant ovarian cancer. *J Clin Oncol* 2014;32(13):1309-16.
- [22] Yardley DA, Noguchi S, Pritchard KI, Burris HA, Baselga J, Gnant M, et al. Everolimus plus exemestane in postmenopausal patients with HR(+) breast cancer: BOLERO-2 final progression-free survival analysis. *Adv Ther* 2013;30(10):870-84.
- [23] Burris 3rd J, Lebrun F, Rugo HS, Beck JT, Piccart M, Neven P, et al. Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. *Cancer* 2013;119(10):1908-15.
- [24] Anota A, Barbieri A, Savina M, Pam A, Gourgou-Bourgade S, Bonnetain F, et al. Comparison of three longitudinal analysis models for the health-related quality of life in

## **Défis dans l'analyse statistique des données de QdV dans les essais cliniques en oncologie**

### **Résumé**

La QdV est reconnue par l'ASCO et la FDA comme un critère de jugement des essais cliniques en oncologie. Dans cet article nous discutons les défis soulevés par l'analyse statistique de la QdV: le schéma d'évaluation de la QdV, l'instrument de mesure de la QdV, la DMCI, la population d'étude, et le type d'analyse longitudinale (modèles mixtes d'analyse de variance sur mesures répétées et TJD).

## Statistical Challenges in the Analysis of Health-Related Quality of Life in Cancer Clinical Trials

Franck Bonnetain, Frédéric Fiteni, Fabio Efficace, and Amélie Anota

Franck Bonnetain, Frédéric Fiteni, and Amélie Anota, University Hospital of Besançon; Franck Bonnetain, Frédéric Fiteni, and Amélie Anota, INSERM U1098, University of Franche-Comté; Franck Bonnetain and Amélie Anota, The French National Platform of Quality of Life and Cancer, Besançon, France; and Fabio Efficace, Italian Group for Adult Hematologic Diseases, Rome, Italy.

Published online ahead of print at [www.jco.org](http://www.jco.org) on April 18, 2016.

Editor's note: Statistics in Brief articles are short communications regarding statistical methods or issues. They are designed to alert and educate the readership about a method or issue that may be unfamiliar to or underused by the clinical research community. Each article will serve as a short primer and may refer the reader to additional sources for detailed information regarding both background and application.

Authors' disclosures of potential conflicts of interest are found in the article online at [www.jco.org](http://www.jco.org). Author contributions are found at the end of this article.

Corresponding author: Franck Bonnetain, PhD, Head of Methodology and quality of life unit in oncology, University Hospital of Besançon, 3 boulevard Fleming, 25030 Besançon, France; e-mail: [fbonnetain@chu-besancon.fr](mailto:fbonnetain@chu-besancon.fr).

© 2016 by American Society of Clinical Oncology

0732-183X/16/3499-1/\$20.00

DOI: 10.1200/JCO.2014.56.7974

Health-related quality of life (HRQOL) in cancer research is critical to more comprehensively evaluate patients' perception of the impact of cancer therapy and their health status on quality of life.<sup>1</sup> American Society of Clinical Oncology acknowledged HRQOL as a key treatment outcome,<sup>2</sup> and it has subsequently been used as a primary or secondary end point across all cancer trials.<sup>3</sup> The goal of this article is to outline key statistical aspects of HRQOL analyses and to highlight future research challenges.<sup>4</sup>

### HOW TO IDEALLY PREPARE HRQOL ANALYSES

The protocol should address key methodological issues<sup>5,6</sup> according to the study design, objectives, and treatment modalities.<sup>7</sup> The optimal schedule of HRQOL assessments (time versus event-driven design) to capture a meaningful difference will guide the longitudinal statistical analysis.

### Identification of Targeted HRQOL Dimensions, Minimal Clinically Important Difference, and Adjustment for Multiple Tests

The prior identification of relevant HRQOL dimensions to target for primary analysis is crucial, in order to limit statistical testing and thereby control false-positive results. Bonferroni's method adjustment can be performed if multiple HRQOL dimensions are selected; however, this method is conservative.<sup>8</sup> The associated minimal clinically important difference (MCID) in HRQOL scores should also be defined.<sup>9</sup>

In the future, the optimal number of dimensions to target and the definition of MCID should be explored according to cancer localization, treatment, and setting. Other adjustment methods for multiple tests should also be investigated.

In a phase IV clinical trial on prostate cancer, Hussain et al<sup>10</sup> targeted five HRQOL dimensions: impotence or erectile dysfunction, libido, vitality, mental health, and physical functioning. They reported that a *P* value of .01 was used to test each

of these dimensions to control the overall type I error rate at .05.

### Study Population

Intent to treat (ITT) is the preferred method to prevent bias, and it enables comparison between studies. Because patients may not have all planned HRQOL measurements, an ITT population subset needs to be defined, ie, all ITT patients with HRQOL data available at baseline (modified ITT population), such as in the AURELIA trial.<sup>11</sup> However, the authors used the term Patient-Reported Outcome population. This reflects the need to harmonize both the definition and the nomenclature of the population data set for HRQOL analysis. Sensitivity analyses could also be recommended to integrate all ITT patients after multiple imputations.

### Missing Data

The validity of HRQOL results depends on both the amount of missing data (MD) and the extent to which these MD are associated with treatment group assignment. Substantial MD reduce statistical power and can introduce a population selection bias if not properly considered in the analysis. Investigation of the mechanisms of MD and sensitivity analyses are thus critical to support the robustness of the results.<sup>12</sup> In cases of informative MD, pattern-mixture models, multiple imputations, or considering patients without HRQOL to be in deterioration should be performed as a sensitivity analysis.<sup>13</sup>

In the AURELIA trial,<sup>11</sup> the reasons for missing HRQOL questionnaires were collected. The main analysis was conducted considering all patients with no evaluable questionnaire at a subsequent time point as not having improved. By sensitivity analysis, only patients whose postbaseline questionnaires were missing as a result of health status were considered to be unimproved.

### HOW TO ANALYZE HRQOL

Some trials only report the mean score and/or the number of patients who deteriorated or improved by treatment arm.<sup>14</sup> Longitudinal analyses

are required to fully examine all collected information and to take into account correlations between repeated measures of HRQOL. Table 1 outlines the advantages and disadvantages of the two main methods for longitudinal analysis of HRQOL data in oncology.

### Linear Mixed Model for Repeated Measures

The linear mixed model (LMM)<sup>15</sup> estimates fixed effects (average trends): a time effect, an arm effect, and a treatment-by-time interaction. Random effects (individual trends) are also introduced. A significant arm effect indicates either a constant difference in HRQOL score over time (from baseline), or a difference on average. A significant treatment-by-time interaction effect indicates that the effect of the treatment on a patient's HRQOL level differs significantly over time. This model accounts for the association (ie, correlation) of repeated measures made on the same patient. Thus, the structure of the covariance matrix between HRQOL measures must be selected according to Akaike's information criteria or Bayesian analysis. The LMM requires a normal distribution of the residuals, which represent the part that is unexplained by the model. Heterogeneity in the choice of effects introduced into the model and in the covariance structure remains challenging.

In the phase III AURELIA (Avastin Use in Platinum-Resistant Epithelial Ovarian Cancer) trial,<sup>11</sup> an LMM model was constructed integrating three fixed effects (arm and time effects, plus a treatment-by-time-interaction) and a random effect on patients. A significant treatment-by-time-interaction of 6.4 points was highlighted for abdominal/gastrointestinal symptoms in favor of the bevacizumab arm ( $P = .015$ ). The results, reported in accordance with published guidelines for HRQOL reporting, including the CONSORT PRO extension,<sup>6,16,17</sup> complement progression-free survival results and confirm the clinical benefit.<sup>18</sup>

### Time-to-Event Longitudinal Analysis

Time-to-event modeling, such as time-to-HRQOL score(s) deterioration (TTD), is also used,<sup>19-23</sup> and can be defined according to the choice of reference score (the MCID), including death (or not) and/or

MD as an event.<sup>24</sup> In the adjuvant setting, the most appropriate definition could be the time-from-inclusion-randomization to a first clinically significant deterioration compared with the baseline score.<sup>25</sup> In the advanced setting, the time until definitive deterioration (TUDD), with or without death as an event, could be more appropriate, reflecting the absorbing state of health deterioration.<sup>19</sup> These models have the advantage of yielding results that can be easily interpreted by clinicians who are accustomed to survival analysis, because the treatment effect can be summarized with a hazard ratio, or restricted means when hazard proportionality is not respected.<sup>26</sup> Longitudinal HRQOL can be described in terms of the median of TUDD or the percentage with or without deterioration at specific time points.

When the TTD is composite (ie, is composed of at least two components, such as deterioration in at least one HRQOL dimension among all dimensions assessed), homogeneous treatment effects on each component and competitive risks should be investigated.<sup>27</sup>

Guidelines to define TTD event(s) by cancer site are required to achieve standardization.<sup>24,28</sup> Kaplan-Meier overestimation of TTD<sup>29</sup> should be addressed by statistical approaches that deal with interval assessment, because the true time of HRQOL deterioration may be unknown.<sup>29</sup> Indeed, in a case of differential intermittent MD, results would be biased toward longer TTD for one arm.

The BOLERO-2 (Everolimus in Combination With Exemestane in the Treatment of Postmenopausal Women With Estrogen Receptor Positive Locally Advanced or Metastatic Breast Cancer Who Are Refractory to Letrozole or Anastrozole) trial, which investigated everolimus plus exemestane in advanced breast cancer, used TUDD with a five-point MCID<sup>19</sup> for longitudinal HRQOL analysis.<sup>21</sup> The median TUDD in global health status was 8.3 versus 5.8 months in favor of everolimus (hazard ratio = 0.74;  $P = .0084$ ). The HRQOL results were consistent with progression-free survival results, thus enhancing the conclusion regarding the clinical benefit of this therapeutic strategy.

Challenges in HRQOL longitudinal analyses include the following key issues: should we select one of these longitudinal statistical approaches, or perform both, to ensure comparisons between trials; and for each of these approaches, how can we improve standardization in statistical modeling and reporting?

Method	Advantages	Disadvantages
Linear mixed model for repeated measures	Well-established approach Robust against missing data	Response shift is not taken into account Not easy for clinicians to understand A sample size cannot easily be defined
Time to quality-of-life score deterioration	Clinically meaningful results Missing data could easily be taken into account, eg, by considering a patient in deterioration just after the last available HRQOL measure, if the corresponding missing data seem to be the result of deterioration in that patient's health status Death can be taken into account Response shift could be taken into account, eg, by considering the best previous score as the reference score A sample size can easily be defined	Needs enough measures Not recommended for dimensions evaluated by a single item  Multiplicity of existing definitions Competitive risk has to be taken into account due to the nature of this time-to-event end point Interval measures have to be taken into account The criterion for deterioration, including MCID, must be defined

Abbreviations: HRQOL, health-related quality of life; MCID, minimal clinically important difference.

THE CHALLENGE OF OUTCOMES INTERPRETATION

Another challenge that deserves further consideration with regard to HRQOL results is how to maximize uptake of findings by clinicians.<sup>30</sup> The ultimate goal of HRQOL analyses is to robustly inform and support clinical decision-making. However, empirical evidence suggests that lack of understanding of HRQOL outcomes may represent a barrier to the use of HRQOL data by oncologists.<sup>31</sup> There is room for improvement in the presentation of HRQOL outcomes to achieve better communication. Future efforts to establish a minimum set of summary statistics that should be reported will contribute to bridging the gap between research and the application of findings in the real-world setting.

HOW TO DEAL WITH THE RESPONSE-SHIFT EFFECT

Response shift is defined as a change in internal standards, values, and/or the conceptualization of HRQOL as a result of health changes.<sup>32</sup> Because of the adaptation to disease, patients may not assess their HRQOL levels with the same internal references during the trial.<sup>33</sup> This can differ according to treatment.<sup>34</sup> For example, a patient's report of lower fatigue after treatment compared with the baseline score could be partially due to a change in the patient's internal reference for assessing fatigue intensity. Such an effect could influence HRQOL results and the interpretation of the true treatment effect. In the same way that the RECIST criteria were adapted (ie, to use best response as the reference instead of baseline), alternative reference scores should be investigated

when response shift occurs to improve analytic interpretation of HRQOL.<sup>24,35,36</sup>

In conclusion, HRQOL is a key end point to assess clinical benefit in cancer trials; however, no guidelines exist for longitudinal analysis, thus precluding comparison between trials. Two phase III trials investigating bevacizumab in glioblastoma assessed HRQOL using the same tools, but with different schedules and statistical approaches (LMM and TTD).<sup>22,37</sup> The divergent results make it difficult to reach any conclusions about the clinical value of adding bevacizumab. The ability of these statistical approaches to capture a treatment effect according to the study design and therapeutic settings should thus be addressed.<sup>38</sup> Other longitudinal models can also be explored, such as item response theory models.<sup>39,40</sup> Finally, the development of guidelines like ongoing SISAQOL and QRECISt is warranted for longitudinal HRQOL analysis and reporting, both to facilitate clinicians' interpretation of HRQOL findings and to allow the use of HRQOL as a coprimary end point with a tumor end point.<sup>41</sup>

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Disclosures provided by the authors are available with this article at [www.jco.org](http://www.jco.org).

AUTHOR CONTRIBUTIONS

**Manuscript writing:** All authors  
**Final approval of manuscript:** All authors

REFERENCES

1. Osoba D: Health-related quality of life and cancer clinical trials. *Ther Adv Med Oncol* 3:57-71, 2011
2. American Society of Clinical Oncology: Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. *J Clin Oncol* 14:671-679, 1996
3. Efficace F, Fayers P, Pusic A, et al: 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* 121:3335-3342, 2015
4. Wilson MK, Collyar D, Chingos DT, et al: Outcomes and endpoints in cancer trials: Bridging the divide. *Lancet Oncol* 16:e43-e52, 2015
5. Calvert M, Blazeby J, Revicki D, et al: Reporting quality of life in clinical trials: A CONSORT extension. *Lancet* 378:1684-1685, 2011
6. Calvert M, Blazeby J, Altman DG, et al: Reporting of patient-reported outcomes in randomized trials: The CONSORT PRO extension. *JAMA* 309:814-822, 2013
7. Fairclough DL: Design and Analysis of Quality of Life Studies in Clinical Trials. Boca Raton, FL, CRC Press, 2010
8. Bland JM, Altman DG: Multiple significance tests: The Bonferroni method. *BMJ* 310:170, 1995
9. Jaeschke R, Singer J, Guyatt GH: Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 10: 407-415, 1989

10. Hussain M, Tangen CM, Berry DL, et al: Intermittent versus continuous androgen deprivation in prostate cancer. *N Engl J Med* 368:1314-1325, 2013
11. Stockler MR, Hilpert F, Friedlander M, et al: Patient-reported outcome results from the open-label phase III AURELIA trial evaluating bevacizumab-containing therapy for platinum-resistant ovarian cancer. *J Clin Oncol* 32:1309-1316, 2014
12. Fairclough DL, Peterson HF, Cella D, et al: Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Stat Med* 17:781-796, 1998
13. Little RJ, Wang Y: Pattern-mixture models for multivariate incomplete data with covariates. *Biometrics* 52:98-111, 1996
14. Fuchs CS, Tomasek J, Yong CJ, et al: Ramucirumab monotherapy for previously treated advanced gastric or gastro-oesophageal junction adenocarcinoma (REGARD): An international, randomised, multicentre, placebo-controlled, phase 3 trial. *Lancet* 383:31-39, 2014
15. Cnaan A, Laird NM, Slasor P: Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Stat Med* 16:2349-2380, 1997
16. Brundage M, Osoba D, Bezjak A, et al: Lessons learned in the assessment of health-related quality of life: Selected examples from the National Cancer Institute of Canada Clinical Trials Group. *J Clin Oncol* 25:5078-5081, 2007
17. Osoba D, Bezjak A, Brundage M, et al: Analysis and interpretation of health-related quality-of-life

data from clinical trials: Basic approach of the National Cancer Institute of Canada Clinical Trials Group. *Eur J Cancer* 41:280-287, 2005

18. Pujade-Lauraine E, Hilpert F, Weber B, et al: Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *J Clin Oncol* 32:1302-1308, 2014
19. Bonnetain F, Dahan L, Maillard E, et al: Time until definitive quality of life score deterioration as a means of longitudinal analysis for treatment trials in patients with metastatic pancreatic adenocarcinoma. *Eur J Cancer* 46:2753-2762, 2010
20. Gourgou-Bourgade S, Bascoul-Mollevis 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* 31:23-29, 2013
21. Burris HA III, Lebrun F, Rugo HS, et al: Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. *Cancer* 119: 1908-1915, 2013
22. Chinot OL, Wick W, Mason W, et al: Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N Engl J Med* 370:709-722, 2014
23. Kabbani FF, Wallace JF, Holmgren E, et al: Health-related quality of life impact of bevacizumab when combined with irinotecan, 5-fluorouracil, and leucovorin or 5-fluorouracil and leucovorin for



metastatic colorectal cancer. *Oncologist* 13:1021-1029, 2008

24. Anota A, Hamidou Z, Paget-Bailly S, et al: Time to health-related quality of life score deterioration as a modality of longitudinal analysis for health-related quality of life studies in oncology: Do we need RECIST for quality of life to achieve standardization? *Qual Life Res* 24:5-18, 2015

25. Hamidou Z, Dabakuyo TS, Mercier M, et al: Time to deterioration in quality of life score as a modality of longitudinal analysis in patients with breast cancer. *Oncologist* 16:1458-1468, 2011

26. Royston P, Parmar MK: 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, 2013

27. Freemantle N, Calvert M, Wood J, et al: Composite outcomes in randomized trials: Greater precision but with greater uncertainty? *JAMA* 289:2554-2559, 2003

28. Bellera CA, Pulido M, Gourgou S, et al: Protocol of the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATE-CAN) project: Formal consensus method for the development of guidelines for standardised time-to-event endpoints' definitions in cancer clinical trials. *Eur J Cancer* 49:769-781, 2013

29. Panageas KS, Ben-Porat L, Dickler MN, et al: When you look matters: The effect of assessment schedule on progression-free survival. *J Natl Cancer Inst* 99:428-432, 2007

30. Brundage M, Bass B, Jolie R, et al: A knowledge translation challenge: Clinical use of quality of life data from cancer clinical trials. *Qual Life Res* 20:979-985, 2011

31. Rouette J, Blazeby J, King M, et al: Integrating health-related quality of life findings from randomized clinical trials into practice: An international study of oncologists' perspectives. *Qual Life Res* 24:1317-1325, 2015

32. Sprangers MA, Schwartz CE: Integrating response shift into health-related quality of life research: A theoretical model. *Soc Sci Med* 48:1507-1515, 1999

33. Schwartz CE, Sprangers MA: Methodological approaches for assessing response shift in longitudinal health-related quality-of-life research. *Soc Sci Med* 48:1531-1548, 1999

34. Ring L, Höfer S, Heuston F, et al: Response shift masks the treatment impact on patient reported outcomes (PROs): The example of individual quality of life in edentulous patients. *Health Qual Life Outcomes* 3:55, 2005

35. Eisenhauer EA, Therasse P, Bogaerts J, et al: New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer* 45:228-247, 2009

36. Therasse P, Arbuck SG, Eisenhauer EA, et al: New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 92:205-216, 2000

37. Gilbert MR, Dignam JJ, Armstrong TS, et al: A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N Engl J Med* 370:699-708, 2014

38. Anota A, Barbieri A, Savina M, et al: Comparison of three longitudinal analysis models for the health-related quality of life in oncology: A simulation study. *Health Qual Life Outcomes* 12:192, 2014

39. Douglas JA: Item response models for longitudinal quality of life data in clinical trials. *Stat Med* 18:2917-2931, 1999

40. Barbieri A, Anota A, Conroy T, et al: Applying longitudinal model from item response theory to assess the health-related quality of life in PRODIGE 4/ ACCORD 11 randomized trial. *Med Decis Making* 0272989X15621883, 2015

41. Fiteni F, Pam A, Anota A, et al: Health-related quality-of-life as co-primary endpoint in randomized clinical trials in oncology. *Expert Rev Anticancer Ther* 15:885-891, 2015

**AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST**

**Statistical Challenges in the Analysis of Health-Related Quality of Life in Cancer Clinical Trials**

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

**Franck Bonnetain**

**Consulting or Advisory Role:** Roche, Celgene, Nestlé, Novartis, Merck Serono, Bristol-Myers Squibb, Eisai, IntegraGen, Invectys

**Research Funding:** Roche (Inst), Novartis (Inst)

**Frédéric Fiteni**

No relationship to disclose

**Fabio Efficace**

**Consulting or Advisory Role:** Bristol-Myers Squibb, Seattle Genetics

**Research Funding:** Bristol-Myers Squibb (Inst), Lundbeck (Inst)

**Amélie Anota**

No relationship to disclose

## **Méthodologie de l'analyse de la qualité de vie relative à la santé dans les essais cliniques de phase III des cancers bronchiques non à petites cellules avancés : une revue systématique**

### **Résumé**

#### **Objectifs**

L'objectif de cette revue était d'évaluer la méthodologie d'analyse et le report de la qualité de vie dans les essais cliniques de phase III des CBNPC avancés.

#### **Méthodes**

Une recherche sur la base de données PubMed a permis d'identifier les essais cliniques de phase III des CBNPC avancés publiés entre Janvier 2008 et décembre 2014. L'extraction a été réalisée par 2 auteurs indépendants.

#### **Résultats**

Cinquante-cinq essais cliniques de phase III de CBNPC avancés ont été identifiés. La QdV était un critère de jugement dans 27 études (49%). Parmi ces 27 études, le QLQ-C30 était utilisé 13 fois (48%) et le FACT-G 12 fois (44%). Les dimensions de QdV cibles, la DMCI, l'approche statistique pour la prise en compte des données manquantes étaient clairement spécifiées dans 13 (48,1%), 9 (33,3%) et 5 (18,5%) études, respectivement. Les méthodes d'analyse statistique les plus fréquemment utilisées étaient : le changement moyen par rapport à baseline (33,3%), le modèle mixte d'analyse de variance sur mesures répétées (22,2%) et le temps jusqu'à détérioration d'un score de QdV (18,5%). Les dimensions cibles, les résultats pour chaque groupe, l'effet traitement et sa précision étaient clairement reportés dans 4 études (14,8%), reportés de manière non claire dans 11 études (40,7%) et non reportés dans 12 études (44,4%).

#### **Conclusions**

Cette revue démontre la faiblesse et l'hétérogénéité de la mesure, analyse et report de la QdV dans les essais cliniques de phase III des CBNPC avancés. Des recommandations précises et uniformes sont nécessaires pour comparer la QdV à travers les publications et transmettre un message compréhensible par les patients et cliniciens.

RESEARCH ARTICLE

Open Access



# Methodology of health-related quality of life analysis in phase III advanced non-small-cell lung cancer clinical trials: a critical review

Frédéric Fiteni<sup>1,2,4\*</sup>, Amélie Anota<sup>1,2,3</sup>, Virginie Westeel<sup>5</sup> and Franck Bonnetain<sup>1,2,3,6</sup>

## Abstract

**Background:** Health-related quality of life (HRQoL) is recognized as a component endpoint for cancer therapy approvals. The aim of this review was to evaluate the methodology of HRQoL analysis and reporting in phase III clinical trials of first-line chemotherapy in advanced non-small cell lung cancers (NSCLC).

**Methods:** A search in MEDLINE databases identified phase III clinical trials in first-line chemotherapy for advanced NSCLC, published between January 2008 to December 2014. Two authors independently extracted information using predefined data abstraction forms.

**Results:** A total of 55 phase III advanced NSCLC trials were identified. HRQoL was declared as an endpoint in 27 studies (49 %). Among these 27 studies, The EORTC questionnaire Quality of Life Questionnaire C30 was used in 13 (48 %) of the studies and The Functional Assessment of Cancer Therapy-General was used in 12 (44 %) trials. The targeted dimensions of HRQoL, the minimal clinically important difference and the statistical approaches for dealing with missing data were clearly specified in 13 (48.1 %), 9 (33.3 %) and 5 (18.5 %) studies, respectively. The most frequent statistical methods for HRQoL analysis were: the mean change from baseline (33.3 %), the linear mixed model for repeated measures (22.2 %) and time to HRQoL score deterioration (18.5 %). For each targeted dimension, the results for each group, the estimated effect size and its precision were clearly reported in 4 studies (14.8 %), not clearly reported in 11 studies (40.7 %) and not reported at all in 12 studies (44.4 %).

**Conclusions:** This review demonstrated the weakness and the heterogeneity of the measurement, analysis, and reporting of HRQoL in phase III advanced NSCLC trials. Precise and uniform recommendations are needed to compare HRQoL results across publications and to provide understandable messages for patients and clinicians.

**Keywords:** Health-related quality of life, Lung cancer, Methodology

## Background

The Food and Drug Administration (FDA) considers overall survival (OS) benefit as the foundation for the approval of new anticancer drugs in the United States [1]. Nevertheless, the increasing number of effective salvage treatments available in many types of cancer (i.e. subsequent lines of treatments) has resulted in the need

for a larger number of patients to be included and/or the need of a more prolonged observation period to attain sufficient events that can achieve planned statistical power; this increases the cost of clinical trials and requires a longer duration to obtain results [2]. Consequently, intermediate endpoints such as progression-free survival are often used as primary endpoints because they are assessed earlier. However, there is a lack of consistency in their definitions [3] and they are not systematically validated as surrogate endpoints for OS.

In this context, HRQoL could constitute an alternative endpoint. HRQoL is recognized as a endpoint for cancer

\* Correspondence: fredericfiteni@gmail.com

<sup>1</sup>Methodology and Quality of Life in Oncology Unit, University Hospital of Besançon, Besançon, France

<sup>2</sup>EA 3181 University of Franche-Comté, Besançon, France

Full list of author information is available at the end of the article



© 2016 Fiteni et al. **Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated.

therapy approvals by the American Society of Clinical Oncology and the FDA [1, 4] HRQoL reflects the patient-perceived evaluation of one's health, including physical, emotional, and social dimensions as well as symptoms due to disease or treatment. Several publications have underlined some key issues regarding the heterogeneity of HRQoL reporting in randomized clinical trials (RCTs) in oncology [5, 6]. Moreover, the statistical longitudinal analysis of HRQoL remains unstandardized which compromises the comparison of results between trials [7]. To improve the reporting of HRQoL in oncology randomized clinical trials, an extension to the CONSORT statement regarding HRQoL reporting was published. Nevertheless, these guidelines do not detail the methodology of statistical analysis of HRQoL. Claassens et al. reported that some aspects of HRQoL reporting (missing HRQoL data, a priori hypotheses of HRQoL, rationale for instruments used) remain underrepresented in non-small-cell lung cancer (NSCLC) RCTs [6].

The aim of this review was to evaluate the three major steps of HRQoL analysis: measurement, statistical analysis, results presentation in phase III clinical trials of first-line chemotherapy in advanced NSCLC with a special focus on the statistical analysis.

## Methods

### Search strategy and Selection for studies

Eligible trials were randomized phase III trials of first-line chemotherapy in advanced NSCLC. Literature searches in PubMed database (January 2008 to December 2014) were performed. Trials published from 2008 were included to assess HRQoL analysis in recent trials. For PubMed database research, the following strategies were used: (lung neoplasm[MeSH Terms]) AND (advanced[Text Word] OR metastatic[Text Word]) + filters Clinical Trial, Phase III. In case of companion papers (i.e. HRQoL analyses reported in a separate paper, not in the princeps one), only the information and methodology declared in the HRQoL paper were reported.

### Data extraction

Two authors (F. F., A.A) independently extracted information using predefined data abstraction forms. All data were checked for internal consistency, and disagreements were resolved by discussion among the investigators. The following details were extracted: general items (number of patients, year of publication, study period, number of centers, nationality of the first author, academic, mixed or industrial trial), name of the primary endpoint, items related to HRQoL measurement and reporting (rationale for HRQoL assessment, methods of data collection, HRQoL questionnaire, evidence of HRQoL questionnaire validity, method/algorithm for scoring the questionnaire, planned schedule of questionnaires

administration, results from each targeted dimensions for multidimensional questionnaires, method for results presentation, discussion of limitations and implications for generalizability and clinical practice.

We assessed statistical HRQoL analysis according to 13 key parameters: statement of the targeted dimensions, statement of the HRQoL hypothesis, procedure to control the type I error rate, statement of the minimal clinically important difference (MCID), data set for HRQoL analyses, description of the number of HRQoL data available at baseline and at subsequent time points, statement of HRQoL scores at baseline for each group and each dimension, profile of missing data at baseline, statement of statistical approaches for dealing with missing data, statistical approach for HRQoL analysis, MCID taken into account in the statistical analysis method and/or in the interpretation of the results, statement of the multivariate analysis. Each key point was coded as "yes" (2 points), "unclear" (1 point) or "no" (0 point). A score on a 0–26 scale was then created with a high score represent a high methodological level for statistical analysis.

### Data analyses

We conducted a descriptive analysis of selected publications, HRQoL measurement, reporting and of the statistical analysis of HRQoL with each key point.

Quantitative variables were described with median and range. Qualitative variables were described with absolute frequencies (number) and relative frequencies (proportion).

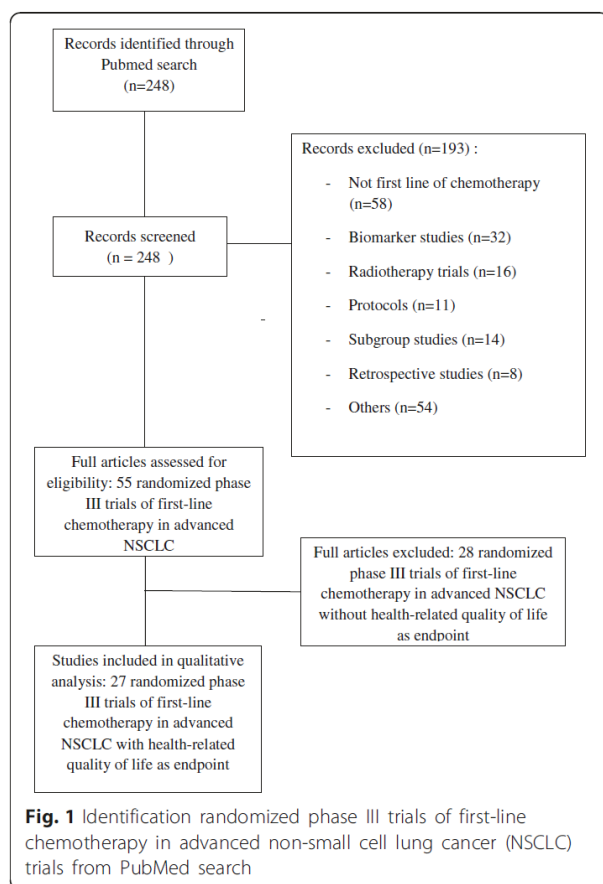
Analyses were conducted with the use of SAS software, version 9.3 (SAS Institute).

## Results

### HRQoL measurement

A total of 55 phase III advanced NSCLC trials published between 2008 and 2014 were identified (Fig. 1). Among these studies, 27 trials were identified with HRQoL endpoint (49 %) (Fig. 1) including 2 studies with HRQoL as primary endpoint. Of the 27 studies, the background and rationale for HRQoL used was provided in 6 trials (22.2 %). For 5 trials (18.5 %), additional HRQoL publications were released in a separate HRQoL dedicated paper. Five trials (18.5 %) provided no result of HRQoL.

The EORTC Quality of Life Questionnaire C30 (QLQ-C30) was the most frequently used instrument. It was used in 13 (48 %) of the studies (Table 1). The lung cancer-specific module EORTC QLQ-LC13 was added to the QLQ-C30 in 12 studies (37 %). Among these 12 studies, two studies added the EuroQoL EQ-5D generic questionnaire to the QLQ-C30 and QLQ-LC13. The Functional Assessment of Cancer Therapy (FACT) questionnaires were used in 12 studies (44.4 %): one study (3.7 %) used the FACT-General (FACT-G), 7 studies



(25.9 %) used the FACT-L questionnaire specific for lung cancer patients and 4 studies (14.8 %) used the FACT-LCS which is a subset of the FACT-L with 7 items (Table 1). The Lung Cancer Symptom Scale questionnaire was used in 1 study (3.7 %). One study did not specify the HRQoL questionnaire used. The reference of the HRQoL instrument validation was provided in 13 studies (48 %). Among the 4 studies using the FACT questionnaires in other than English version, none of them reported the transcultural validation. Six studies (22.2 %) defined the procedure of questionnaire completion (i.e. paper and pencil, electronic completion, at home or at the clinic). The planned schedule of HRQoL assessment was reported in 23 trials (85.2 %).

#### Statistical analysis of HRQoL

The Table 2 summarized the quality of statistical analysis of HRQoL.

The mean score based on the 13 key parameters of the 27 trials was 6.3 (standard deviation = 6.1, range = 0-20).

The targeted dimensions of HRQoL were pre-specified in 13 studies (48.1 %) in the method section: 6 of them used EORTC questionnaires and 7 used

FACT questionnaires. Among the 6 studies which used EORTC questionnaires, 3 dimensions were targeted in mean and targeted dimensions were dyspnea (66.6 %), global health status (66.6 %), pain (50.0 %), fatigue (50.0 %), cough (33.3 %) and nausea/vomiting (33.3 %) (Table 1). Among the 7 studies which used FACT questionnaires, 2 dimensions were targeted in mean and most frequent targeted dimensions were the lung cancer subscale (57.2 %), the trial outcome index (57.2 %) and the FACT-L global score (28.6 %) (Table 1).

Nine studies (33.3 %) defined the MCID. Among these studies: 5 used the EORTC questionnaires and all of them used a 10-point decrease in the HRQoL scores as the MCID (Table 1), 4 used the FACT questionnaires and all of them used 6 points decrease for the FACT-L global score and trial outcome index, and 2 points decrease for the lung cancer subscale as the MCID.

The MCID was taken into account in the statistical analysis and/or in the interpretation of the results in 7 studies (25.9 %) (Table 1).

Only three studies mentioned the population data set for HRQoL analysis: 2 in modified intention-to-treat, 1 in intention-to-treat. Definition for the mITT population were 1) "all randomly assigned patients with data were included" [8] and 2) "patients with a baseline and at least one post-baseline HRQoL assessment were included" [9].

The number of HRQoL data at baseline and at subsequent time points, the HRQoL scores at baseline for each group and each dimension, the profile of missing data at baseline, the statistical approaches for dealing with missing data were adequately reported in 7 (25.9 %), 6 (22.2 %), 1 (3.7 %) and 5 (18.5 %) studies, respectively (Table 2). The statistical methods for dealing with missing data were different (Table 1). No study provided the reasons why data were missing.

Fourteen studies (51.9 %) described the statistical approach to analyze HRQoL data in the method section (Table 2). Seven studies used two or more statistical approaches. The different statistical methods/analyses were: the mean change from baseline (33.3 %), the linear mixed model for repeated measures (LMM) (22.2 %), time to HRQoL score deterioration (TTD) (18.5 %), AUC (7.4 %), mixed-effects growth-curve model (7.4 %), distribution of patients whose symptom had improved, remained stable or worsened (3.4 %), Fisher test for equal proportion of patients achieving at least two points increase (3.4 %), group comparisons of scale at each time (3.4 %), rates of symptom palliation (3.4 %), percentage of patients with at least two points improvement at the beginning of the cycle two (3.4 %) (Table 1). No study specified which were the primary statistical analysis and the sensitive analysis.

Among the six studies which used a LMM to analyze the longitudinal HRQoL data, none of them declared the

**Table 1** Aspects relevant to HRQoL statistical analysis

First authors	HRQoL questionnaire	Timing of assessment	Targeted dimensions	MCID	Statistical approach for dealing with missing data	Statistical approach for HRQoL analysis
Wu [19]	QLQ-C30; QLQ-LC13	Randomization and every 3 weeks until disease progression or new cancer treatment	Cough, dyspnoea, pain	10 points		Distribution of patients whose symptom had improved, remained stable, or worsened; the time to deterioration of symptoms; mixed-effects growth curve model
Laurie [20]	QLQ-C30; QLQ-LC13 and two additional questions (hand-foot syndrome and headache)	At baseline and every cycle				
Yang [7]	QLQ-C30; QLQ-LC13	Random assignment and every 21 days until disease progression	Cough, dyspnoea, pain	10	Joint analysis	Time to deterioration; mixed-effects growth curve model
Wu [19]	FACT-L	Not available	FACT-L total score and TOI			Time to deterioration
Yoshioka [21]	FACT-L; FACT/GOGNTX	At the time of enrollment and at 6 and 9 weeks after the initiation of treatment				Linear mixed model for repeated measures
Lee [22]	QLQ-C30, QLQ-LC14, EUROQOL	Baseline, monthly during the first year, then 18 and 24 months after randomization				Linear mixed model for repeated measures
Gridelli [23]						
Flotten [24]	QLQ-C30, QLQ-LC13	Before each cycle, 3 weeks after the last cycle, and then every 8 weeks until 57 weeks	Global health status health, nausea/vomiting, dyspnoea, fatigue			Mean change from baseline
Socinsky [25]	FACT-G, FACTTAXANE	Baseline and on day 1 of each cycle	Peripheral neuropathy, pain, hearing, edema			Mean change from baseline
Groen [26]	QLQ-C30, QLQ-LC13	Start of chemotherapy and weeks 6,12,16,24,30				Mean change from baseline
Chen [9]	FACT-L	Baseline and every 6 weeks				Mean change from baseline
Lara [27]	QLQ-C30	On day 1 of each odd cycle and at the end of the treatment visit				Mean change from baseline
Koch [12]	QLQ-C30, QLQ-LC13	Baseline, weeks 3,6,9,12,20,28	Dyspnoea, fatigue, pain, pain medication, global health status			Group comparisons of scale at each time; mean change from baseline; AUC; rates of symptom palliation
Biesma [28]	QLQ-C30, QLQ-LC13	Before, during (day 1 and day 8 of each cycle) and after each cycle (weeks 12,15,18)	Global health status	10		Mean change from baseline and week 18; linear mixed model for repeated measures

**Table 1** Aspects relevant to HRQoL statistical analysis (Continued)

Weissman [29]	FACT-L	After each cycle	TOI			Mean change from baseline and after 6 cycles
Okamoto [30]	FACT-L, FACT/GOGNTX	Time of enrollment and at 6 and 9 weeks after initiation of treatment				Linear mixed model for repeated measures
Thongprasert [31]	FACT-L	Baseline, weeks 1 AND 3, 3-weekly until week 18, 6 weekly until progression and at discontinuation	FACT-L Score, TOI, LCS	6 for FACT-L total score and TOI; 2 points for LCS	If less than 50% of the Fact-I subscale scores were missing, the subscale score was divided by the number of completed items and multiplied by the total number of items on the scale. If 50% or more of the items were missing, that subscale was treated as missing for that patient	Mean change from baseline; Time to worsening; Time to improvement
Lynch [32]	FACT-L	Baseline, before each treatment cycle and at the end of the therapy				Mean change from baseline; Time to symptomatic disease progression
Takeda [33]	FACT-L	The time of enrollment and at 12 weeks, 18 weeks after initiation of treatment	LCS		Linear mixed-effects model in which the missing data depend on the observed score	Linear mixed model for repeated measures
Lee [34]	QLQ-C30, QLQ-LC14	Random, During each cycle, at the end of the chemotherapy, every 6 months until 24 months				Linear mixed model for repeated measures
Treat [35]	FACT-L	Not available				
Tan [36]	Lung cancer symptom scale	Baseline, at the end of each cycle, just before the next cycle, at the end of the study				
Pirker [37]	QLQ-C30, QLQ-LC13, EuroQoL EQ-5D	Not available				
Gronberg [38]	QLQ-C30, QLQ-LC13	Weeks 0,3,6,9,12,20,28,36,44,52	Global health status health, nausea/vomiting, dyspnoea, fatigue	10	Last value carried forward for missing value that followed, even after death	Area under the curve
O'Brien [39]	FACT-L	Baseline, before each cycle and at the end of the treatment	LCS	2	Missing scores at week 3 were classified as having less than a 2-point increase in the primary analysis data but classified as missing and excluded from the supplemental analysis	Fisher test for equal proportion of patients achieving at least two points increase
Langer [40]	FACT-L	Baseline and within 3 days of each treatment	LCS	2		Percentage of patients with at least two points improvement at the beginning of cycle 2
Gebbia [41]	QLQ-C30; QLQ-LC13	Baseline and every cycle				



**Table 2** The 13 keys parameters for statistical HRQoL analysis assessed as “yes” if the authors specified the parameter, “not clear” it was not clear and “no” if the authors didn’t specify the parameter

	Yes, n (%)	Not clear, n (%)	No, n (%)
Targeted dimensions	13 (48.1)	0	14 (51.9)
HRQoL hypothesis	2 (7.4)	0	25 (92.6)
Procedure to control the type I error	1 (3.7)	1 (3.7)	25 (92.6)
Minimal clinically important difference	9 (33.3)	1 (3.7)	17 (63)
Study population	3 (11.1)	3 (11.1)	21 (77.8)
Number of HRQoL data at subsequent time points	7 (25.9)	6 (22.2)	14 (51.9)
HRQoL scores at baseline for each group and each dimension	6 (22.2)	2 (7.4)	19 (70.4)
Profile of missing data at baseline	1 (3.7)	2 (7.4)	24 (88.9)
Statistical approaches for dealing with missing data	5 (18.5)	0	22 (81.5)
Statistical approach for HRQoL analysis	14 (51.9)	1 (3.7)	12 (44.4)
MCID taken into account in the statistical analysis	7 (25.9)	2 (7.4)	18 (66.7)
Multivariate analysis	1 (3.7)	0	26 (96.4)

effects introduced in the model (random or fixed effects and interactions).

One study presented a multivariate analysis (Table 2).

#### HRQoL results presentation

For each targeted dimension, the results for each group, the estimated effect size and its precision were adequately reported in four studies (14.8 %), not clearly reported in 11 studies (40.7 %) and not reported at all in 12 studies (44.4 %). A discussion specific limitations and implications for clinical practice was provided in 10 articles (37 %) (Table 1).

The results were presented in the text, figures or tables in 24 (88.9 %), 12 (44.4 %) and 8 (29.6 %) articles, respectively (Table 1).

#### Discussion

In this review, we showed that HRQoL was declared as an endpoint in only 49 % of the phase III clinical trials in advanced NSCLC published between 2008 and 2014. Moreover, we clearly demonstrated the heterogeneity and the weakness of the methodology of HRQoL measurement, statistical analysis and reporting. First, two questionnaires are the most widely used: the QLQ-C30 plus LC13 module (48 %) and the FACT-L (44 %). As an example, the OPTIMAL [8, 9] and the Lux-Lung 3 [10] trials compared the efficacy and tolerability of two tyrosine kinase inhibitor (erlotinib and afatinib, respectively) versus chemotherapy in first-line line treatment of patients with advanced EGFR mutation-positive NSCLC. HRQoL was assessed by the FACT-L questionnaire in OPTIMAL trial [9] while EORTC QLQ-C30 and LC13 module were used in LUX-Lung 3 trial [7]. These two clinical trials could hardly be compared since EORTC and FACT questionnaires for lung cancer do not contain

the same dimensions in regard to impact of lung cancer on HRQoL. Moreover, two studies used the EuroQoL EQ-5D generic questionnaire which comprises only five short questions and is suitable since it limits patient burden and in that way also encourages response rate. Nevertheless this questionnaire has not been tailored to the special requirements of patients with cancer. At this time, the foremost challenge would be to promote, through cancer site and treatment modalities, guidelines for selecting the best questionnaires allowing for direct comparison of results across trials. Moreover, it is also still necessary to develop some new tools to evaluate HRQoL. Already validated questionnaires may not be adapted to new targeted biotherapy agent which can induce some long-term moderate toxicities.

Then, the number of HRQoL measures and intervals between two consecutive measures vary from one study to another. HRQoL is often captured until tumor progression, nevertheless, in advanced NSCLC, we could wonder if it would be more appropriate to measure HRQoL until death. Recommendations on the schedule of HRQoL assessment should be provided. At least three HRQoL assessment are recommended: at baseline, during treatment and at the end of the study (<http://group-s.eortc.be/qol/eortc-qlq-c30>). However, a more intensive HRQoL assessment is preferable to better capture the longitudinal trajectory of HRQoL level and to capture any relevant changes. In this review, most of studies evaluate HRQoL at baseline and then every treatment cycle which allow a good appreciation of the impact of treatment of HRQoL over time, but it depends on the number of cycles received by the patient. Moreover attention should be paid to the timing of diagnostic procedures which could influenced HRQoL results, especially with lifethreatening cancers. Patients are likely to be

experiencing stress in anticipation of the yet unknown results. After the procedure, the patients will either be experiencing great relief or anxiety depending on the results. This point should be taken into account in HRQoL assessment design and be carefully documented in the protocol and emphasized during training.

The *a priori* selection of the targeted dimensions was heterogeneous between the trials and was pre-specified only 13 times (48.1 %) in the method section and most of them are symptomatic scales. We know that there is general agreement concerning the multidimensional concept of HRQoL taking into account levels of physical, mental, social, and patient satisfaction with treatment. Therefore, the choice of only symptomatic HRQoL dimensions reaches the problem of the holistic sense of HRQoL. Moreover, the choice of the targeted dimensions of HRQoL must be discussed between clinicians and methodologists and clearly described in the protocol.

In confirmatory clinical trials with multiple endpoints, the use of multiple test procedures is mandatory and CONSORT Statement recommends a multiplicity adjustment in case of multiple testing [11]. However, only one study clearly stated the procedure to control the type I error [12].

Prior to longitudinal HRQoL data analysis, the MCID should be *a priori* determined [13]. In our review, only nine studies (33.3 %) clearly specified it. The MCID represents the smallest changes/differences in HRQoL score, which is perceived as clinically important. For the EORTC questionnaires, a 5-point to a 10-point difference in scores could be considered as the MCID [14]. In patients with NSCLC, Maringwa et al. [15] tried to determine the smallest changes in HRQoL scores in a subset of the EORTC QLQ-C30 scales, which could be considered as clinically meaningful. They concluded that the estimates of 5 to 10 units of the QLQ-C30 scales may be used as guidance for clinicians and researchers to classify patients as improved or deteriorated. In our review, the 5 studies which used the QLQ-C30 and stated the MCID, all used a 10-point decrease in the HRQoL scores as the MCID. A sensitivity analysis, with a MCID of 5 points could have been proposed to assess the robustness of the results.

Missing data, considered as missing not at random, can bias the longitudinal analysis if it is not adequately taken into account [16]. Patients may drop out before the planned end of the study, resulting in the absence of any available HRQoL data after the patient's drop out (i.e. attrition). Moreover, drop out occurs generally due to a deterioration of patient health status or death. Patients may also be too tired to fill the questionnaire entirely at a specific measurement time. This induces the potential risk to select subpopulation of patients with better HRQoL levels and with available HRQoL data.

Not adjusting for missing data can limit the robustness of the results and the confidence in the HRQoL conclusions. Therefore, the profile of missing data at baseline and the number of HRQoL at subsequent time points for each group must be specified. In our review, these two information were reported in only 1 (3.7 %) and 7 (25.9 %) trials, respectively. Furthermore, only 5 studies (18.5 %) described the statistical approaches for dealing with missing data and the 5 methods were different. Thus, the reporting of missing data and the statistical approaches of analysis of missing data need to be standardized.

There are a number of possible ways of analyzing longitudinal HRQoL data. Currently, the two main robust methods are: the LMM and the TTD [6, 17]. In our review, the most widely used is the mean change from baseline (33.3 %). This method summarizes the longitudinal data into a summary statistic before performing a between-arms comparison. This method is rather simple but may overlook important changes in HRQoL along time and should not be applied. This method is not a model of longitudinal analysis. The LMM method was used in 6 studies (22.2 %). In all these studies the fixed effects, the random effects and the correlation matrix between HRQoL measures introduced in the model were not specified. Thus, results are very difficult to interpret. The LMM can estimate a time effect, an arm effect and an interaction between treatment arm and time (reflecting a different evolution of the two treatment arms over time). The LMM contains both fixed effects (reflecting average trends) such as treatment and random effects (individual trends). This model accounts for the association (i.e. correlation) of measures made on the same patient at different times.

Finally, the LMM model generally require a normally distribution of the score studied. All studies using this method did not mention this hypothesis and *a priori* did not check it. For EORTC questionnaires, scores generally do not respect a normal distribution due to the low number of items per dimension. The TTD approach was used in 5 studies (18.5 %). Currently the definition of the TTD is not standardized; therefore the studies must clearly define it. The definition of the TTD must include: reference score, the event of interest, the censoring process, including death or not [6]. In our review, only one study defined the TTD: "Time to deterioration in patient-reported outcomes was measured in months from random assignment to the first instance of symptom worsening (10 points from baseline). Patients without worsening, including those with disease progression, were censored at the last available patient-reported outcome assessment; those lacking post-baseline assessments were censored at random assignment. Patients who died without documented worsening were considered to have

deteriorated at the time of death” [7]. Other statistical methods were used: “distribution of patients whose symptom had improved, remained stable or worsened”, “Fisher test for equal proportion of patients achieving at least two points increase”, “group comparisons of scale at each time”, “rates of symptom palliation”, “percentage of patients with at least two points improvement at the beginning of the cycle two”. These methods can guide on the choice of the most appropriate analytical method to use to analyze longitudinal HRQoL data (e.g., the LMM or the TTD) approach. However, these analyses alone cannot be sufficient to capture all information present in the longitudinal HRQoL assessment and must be completed by a longitudinal statistical approach taking into account the correlation between HRQoL measures.

Finally, it should be acknowledged that journal space is often limited, and authors may not have been able to report all the methodology. Therefore, HRQoL may systematically be reported in separate HRQoL dedicated manuscripts.

## Conclusions

Our review demonstrated the poor quality and the heterogeneity of the measurement, analysis, and reporting of HRQoL in phase III advanced NSCLC trials. The heterogeneity between trials limits their cross comparison and the feasibility of meta-analysis.

The HRQoL CONSORT statement regarding HRQoL reporting was published in 2013 [18] and it is true that we reviewed studies published before 2013. The use of the HRQoL CONSORT extensions should be encouraged. Nevertheless, these guidelines do not detail the methodology of statistical analysis of HRQoL. Incomplete or inaccurate statistical analysis of HRQoL can affect the reliability of these outcomes. Therefore, development of guidelines for longitudinal HRQoL analysis of clinical trials is important to facilitate interpretation of HRQoL findings.

## Abbreviations

HRQoL: Health-related quality of life; NSCLC: Non-small cell lung cancer; FDA: Food and Drug Administration; OS: Overall survival; RCT: Randomized clinical trial; MCID: Minimal clinically important difference; EORTC QLQ-C30: European organization for research and treatment of cancer quality of life questionnaire C30; FACT: Functional Assessment of Cancer Therapy; LMM: Linear mixed model for repeated measures; TTD: Time to health-related quality of life score deterioration.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contribution

Conception and design: FF, AA, VW, FB. Acquisition of data: FF, AA, Analysis and interpretation of data: FF, AA, VW, FB. Involved in drafting the manuscript or revising it critically for important intellectual content: FF, AA, VW, FB. Agree to be accountable for all aspects of the work in ensuring that questions related to the Accuracy or integrity of any part of the work are appropriately investigated and resolved: FF, AA, VW, FB. All authors have read and approved the manuscript.

## Acknowledgements

No acknowledgement.

## Author details

<sup>1</sup>Methodology and Quality of Life in Oncology Unit, University Hospital of Besançon, Besançon, France. <sup>2</sup>EA 3181 University of Franche-Comté, Besançon, France. <sup>3</sup>The French National Platform Quality of Life and Cancer, Besançon, France. <sup>4</sup>Department of Medical Oncology, University Hospital of Besançon, Besançon, France. <sup>5</sup>Chest disease Department, University Hospital of Besançon, Besançon, France. <sup>6</sup>EORTC QOL Group, Brussels, Belgium.

Received: 1 July 2015 Accepted: 9 February 2016

Published online: 18 February 2016

## References

- Beitz J, Gnecco C, Justice R. Quality-of-life end points in cancer clinical trials: the U.S. Food and Drug Administration perspective. *J Natl Cancer Inst Monogr.* 1996;(20):7-9. <http://www.fda.gov/downloads/Drugs/.../Guidances/ucm071590.pdf>
- Fiteni F, Westeel V, Pivot X, Borg C, Vernerey D, Bonnetain F. Endpoints in cancer clinical trials. *J Visc Surg.* 2014;151(1):17-22.
- Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J Clin Oncol Off J Am Soc Clin Oncol.* 2008;26(22):3721-6.
- Johnson JR, Temple R. Food and Drug Administration requirements for approval of new anticancer drugs. *Cancer Treat Rep.* 1985;69(10):1155-9.
- Claessens L, van Meerbeeck J, Coens C, Quinten C, Ghislain I, Sloan EK, et al. Health-related quality of life in non-small-cell lung cancer: an update of a systematic review on methodologic issues in randomized controlled trials. *J Clin Oncol Off J Am Soc Clin Oncol.* 2011;29(15):2104-20.
- Anota A, Hamidou Z, Paget-Bailly S, Chibaudel B, Bascoul-Mollevi C, Auquier P, 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 Int J Qual Life Asp Treat Care Rehabil.* 2015;24(1):5-18.
- Yang JC-H, Hirsh V, Schuler M, Yamamoto N, O'Byrne KJ, Mok TSK, et al. Symptom control and quality of life in LUX-Lung 3: a phase III study of afatinib or cisplatin/pemetrexed in patients with advanced lung adenocarcinoma with EGFR mutations. *J Clin Oncol Off J Am Soc Clin Oncol.* 2013;31(27):3342-50.
- Zhou C, Wu Y-L, Chen G, Feng J, Liu X-Q, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol.* 2011;12(8):735-42.
- Chen G, Feng J, Zhou C, Wu Y-L, Liu X-Q, Wang C, et al. Quality of life (QoL) analyses from OPTIMAL (CTONG-0802), a phase III, randomised, open-label study of first-line erlotinib versus chemotherapy in patients with advanced EGFR mutation-positive non-small-cell lung cancer (NSCLC). *Ann Oncol Off J Eur Soc Med Oncol ESMO.* 2013;24(6):1615-22.
- Sequist LV, Yang JC-H, Yamamoto N, O'Byrne K, Hirsh V, Mok T, et al. Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. *J Clin Oncol Off J Am Soc Clin Oncol.* 2013;31(27):3327-34.
- Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg Lond Engl.* 2012;10(1):28-55.
- Koch A, Bergman B, Holmberg E, Sederholm C, Ek L, Kosieradzki J, et al. Effect of celecoxib on survival in patients with advanced non-small cell lung cancer: a double blind randomised clinical phase III trial (CYCLUS study) by the Swedish Lung Cancer Study Group. *Eur J Cancer Oxf Engl.* 1990. 2011;47(10):1546-55.
- Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertain the minimal clinically important difference. *Control Clin Trials.* 1989;10(4):407-15.
- 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):139-44.
- Maringwa JT, Quinten C, King M, Ringash J, Osoba D, Coens C, et al. Minimal important differences for interpreting health-related quality of life

- scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials. *Support Care Cancer Off J Multinatl Assoc Support Care Cancer*. 2011;19(11):1753–60.
16. Bernhard J, Cella DF, Coates AS, Fallowfield L, Ganz PA, Moynour CM, et al. Missing quality of life data in cancer clinical trials: serious problems and challenges. *Stat Med*. 1998;17(5-7):517–32.
  17. Fiteni F, Anota A, Westeel V, Bonnetain F. [Health-related quality of life in phase III cancer clinical trials: From questionnaire administration to statistical analysis]. *Bull Cancer (Paris)*. 2015;102(4):360–6.
  18. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA J Am Med Assoc*. 2013;309(8):814–22.
  19. Wu Y-L, Lee JS, Thongprasert S, Yu C-J, Zhang L, Ladrera G, et al. Intercalated combination of chemotherapy and erlotinib for patients with advanced stage non-small-cell lung cancer (FASTACT-2): a randomised, double-blind trial. *Lancet Oncol*. 2013;14(8):777–86.
  20. Laurie SA, Solomon BJ, Seymour L, Ellis PM, Goss GD, Shepherd FA, et al. Randomised, double-blind trial of carboplatin and paclitaxel with daily oral cediranib or placebo in patients with advanced non-small cell lung cancer: NCIC Clinical Trials Group study BR29. *Eur J Cancer Oxf Engl*. 1990. 2014; 50(4):706–12.
  21. Yoshioka H, Okamoto I, Morita S, Ando M, Takeda K, Seto T, et al. Efficacy and safety analysis according to histology for S-1 in combination with carboplatin as first-line chemotherapy in patients with advanced non-small-cell lung cancer: updated results of the West Japan Oncology Group LETS study. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. 2013;24(5):1326–31.
  22. Lee SM, Khan I, Upadhyay S, Lewanski C, Falk S, Skailes G, et al. First-line erlotinib in patients with advanced non-small-cell lung cancer unsuitable for chemotherapy (TOPICAL): a double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol*. 2012;13(11):1161–70.
  23. Gridelli C, Ciardiello F, Gallo C, Feld R, Butts C, Gebbia V, et al. First-line erlotinib followed by second-line cisplatin-gemcitabine chemotherapy in advanced non-small-cell lung cancer: the TORCH randomized trial. *J Clin Oncol Off J Am Soc Clin Oncol*. 2012;30(24):3002–11.
  24. Fløtten Ø, Grønberg BH, Bremnes R, Amundsen T, Sundstrøm S, Rolke H, et al. Vinorelbine and gemcitabine vs vinorelbine and carboplatin as first-line treatment of advanced NSCLC. A phase III randomised controlled trial by the Norwegian Lung Cancer Study Group. *Br J Cancer*. 2012;107(3):442–7.
  25. Socinski MA, Bondarenko I, Karaseva NA, Makhson AM, Vynnychenko I, Okamoto I, et al. Weekly nab-paclitaxel in combination with carboplatin versus solvent-based paclitaxel plus carboplatin as first-line therapy in patients with advanced non-small-cell lung cancer: final results of a phase III trial. *J Clin Oncol Off J Am Soc Clin Oncol*. 2012;30(17):2055–62.
  26. Groen HJM, Sietsma H, Vincent A, Hochstenbag MMH, van Putten JWG, van den Berg A, et al. Randomized, placebo-controlled phase III study of docetaxel plus carboplatin with celecoxib and cyclooxygenase-2 expression as a biomarker for patients with advanced non-small-cell lung cancer: the NVALT-4 study. *J Clin Oncol Off J Am Soc Clin Oncol*. 2011;29(32):4320–6.
  27. Lara PN, Douillard J-Y, Nakagawa K, von Pawel J, McKeage MJ, Albert I, et al. Randomized phase III placebo-controlled trial of carboplatin and paclitaxel with or without the vascular disrupting agent vandimezan (ASA404) in advanced non-small-cell lung cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2011;29(22):2965–71.
  28. Biesma B, Wymenga ANM, Vincent A, Dalesio O, Smit HJM, Stigt JA, et al. Quality of life, geriatric assessment and survival in elderly patients with non-small-cell lung cancer treated with carboplatin-gemcitabine or carboplatin-paclitaxel: NVALT-3 a phase III study. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. 2011;22(7):1520–7.
  29. Weissman CH, Reynolds CH, Neubauer MA, Pritchard S, Kobina S, Asmar L. A phase III randomized trial of gemcitabine-oxaliplatin versus carboplatin-paclitaxel as first-line therapy in patients with advanced non-small cell lung cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2011;6(2):358–64.
  30. Okamoto I, Yoshioka H, Morita S, Ando M, Takeda K, Seto T, et al. Phase III trial comparing oral S-1 plus carboplatin with paclitaxel plus carboplatin in chemotherapy-naïve patients with advanced non-small-cell lung cancer: results of a west Japan oncology group study. *J Clin Oncol Off J Am Soc Clin Oncol*. 2010;28(36):5240–6.
  31. Thongprasert S, Duffield E, Saijo N, Wu Y-L, Yang JC-H, Chu D-T, et al. Health-related quality-of-life in a randomized phase III first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients from Asia with advanced NSCLC (IPASS). *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2011;6(11):1872–80.
  32. Lynch TJ, Patel T, Dreisbach L, McCleod M, Heim WJ, Hermann RC, et al. Cetuximab and first-line taxane/carboplatin chemotherapy in advanced non-small-cell lung cancer: results of the randomized multicenter phase III trial BMS099. *J Clin Oncol Off J Am Soc Clin Oncol*. 2010;28(6):911–7.
  33. Takeda K, Hida T, Sato T, Ando M, Seto T, Satouchi M, et al. Randomized phase III trial of platinum-doublet chemotherapy followed by gefitinib compared with continued platinum-doublet chemotherapy in Japanese patients with advanced non-small-cell lung cancer: results of a west Japan thoracic oncology group trial (WJTOG0203). *J Clin Oncol Off J Am Soc Clin Oncol*. 2010;28(5):753–60.
  34. Lee SM, Rudd R, Woll PJ, Ottensmeier C, Gilligan D, Price A, et al. Randomized double-blind placebo-controlled trial of thalidomide in combination with gemcitabine and Carboplatin in advanced non-small-cell lung cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2009;27(31):5248–54.
  35. Treat JA, Gonin R, Socinski MA, Edelman MJ, Catalano RB, Marinucci DM, et al. A randomized, phase III multicenter trial of gemcitabine in combination with carboplatin or paclitaxel versus paclitaxel plus carboplatin in patients with advanced or metastatic non-small-cell lung cancer. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. 2010;21(3):540–7.
  36. Tan EH, Rolski J, Grodzki T, Schneider CP, Gatzemeier U, Zatloukal P, et al. Global Lung Oncology Branch trial 3 (GLOB3): final results of a randomised multinational phase III study alternating oral and i.v. vinorelbine plus cisplatin versus docetaxel plus cisplatin as first-line treatment of advanced non-small-cell lung cancer. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. 2009;20(7):1249–56.
  37. Pirker R, Pereira JR, Szczesna A, von Pawel J, Krzakowski M, Ramlau R, et al. Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. *Lancet*. 2009;373(9674):1525–31.
  38. Grønberg BH, Bremnes RM, Fløtten Ø, Amundsen T, Brunsvig PF, Hjelde HH, et al. Phase III study by the Norwegian lung cancer study group: pemetrexed plus carboplatin compared with gemcitabine plus carboplatin as first-line chemotherapy in advanced non-small-cell lung cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 2009;27(19):3217–24.
  39. O'Brien MER, Socinski MA, Popovich AY, Bondarenko IN, Tomova A, Bilynsky BT, et al. Randomized phase III trial comparing single-agent paclitaxel Poliglumex (CT-2103, PPX) with single-agent gemcitabine or vinorelbine for the treatment of PS 2 patients with chemotherapy-naïve advanced non-small cell lung cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2008;3(7):728–34.
  40. Langer CJ, O'Byrne KJ, Socinski MA, Mikhailov SM, Leśniewski-Kmak K, Smakal M, et al. Phase III trial comparing paclitaxel poliglumex (CT-2103, PPX) in combination with carboplatin versus standard paclitaxel and carboplatin in the treatment of PS 2 patients with chemotherapy-naïve advanced non-small cell lung cancer. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2008;3(6):623–30.
  41. Gebbia V, Galetta D, Lorusso V, Caruso M, Verderame F, Pezzella G, et al. Cisplatin plus weekly vinorelbine versus cisplatin plus vinorelbine on days 1 and 8 in advanced non-small cell lung cancer: a prospective randomized phase III trial of the G.O.I.M. (Gruppo Oncologico Italia Meridionale). *Lung Cancer Amst Neth*. 2008;61(3):369–77.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)



# **La qualité de vie relative à la santé chez les patients âgés atteints d'un carcinome bronchique non à petites cellules avancé traités par le doublet de chimiothérapie carboplatine et paclitaxel hebdomadaire en comparaison à une monochimiothérapie**

## **Résumé**

### **Contexte**

L'étude de phase III *Intergroupe Francophone de Cancérologie Thoracique* (IFCT) 0501 a démontré un bénéfice significatif sur la survie globale mais avec une augmentation de la toxicité dans le bras carboplatine-paclitaxel en comparaison à une monochimiothérapie chez les patients âgés atteints d'un carcinome bronchique non à petites cellules avancé. Nous avons comparé longitudinalement la QdV entre les 2 bras de traitement selon la technique du temps jusqu'à détérioration définitive d'un score de QdV.

### **Méthodes**

Au total, 451 patients âgés de 70-89 ans atteints de CBNPC avancés ont été randomisés entre 4 cycles de carboplatine plus paclitaxel hebdomadaire et 5 cycles de vinorelbine ou gemcitabine. La QdV a été mesurée par le questionnaire QLQ-C30 à baseline, 6 semaines et 18 semaines. Le TJD a été défini par le temps entre la randomisation et la première décroissance d'un score du QLQ-C30  $\geq 5$  points sans amélioration ultérieure d'un score du QLQ-C30  $\geq 5$  points ou autre donnée de QdV disponible.

### **Résultats**

En utilisant une DMCI de 5 points, les patients traités par le doublet de chimiothérapie avaient un TJD supérieur pour 2 dimensions de QdV: fonctionnement physique (médiane TJD: 2.04 pour le doublet *vs.* 1.71 months pour la monothérapie; log-rank  $p=0.01$ ) et nausées/vomissements (median: non atteinte *vs.* 4.83, respectivement; log-rank  $p=0.046$ ). Le modèle de Cox multivarié a montré que le bras carboplatine/paclitaxel était indépendamment associé à un TJD supérieur pour ces 2 dimensions de QdV. De plus, il n'y avait de différence significative en TJD entre les 2 bras pour les autres dimensions de QdV.

### **Conclusion**

Malgré une augmentation de la toxicité, le doublet de chimiothérapie ne diminue pas le TJD chez les patients âgés atteints de CBNPC avancés. De plus, l'augmentation de la SG pour le doublet de chimiothérapie est associée à une augmentation du TJD pour 2 dimensions de QdV, à savoir le fonctionnement physique et les nausées/vomissements.



# Health-related quality of life in elderly patients with advanced non-small cell lung cancer comparing carboplatin and weekly paclitaxel doublet chemotherapy with monotherapy

Frédéric Fiteni<sup>1,2,3</sup>, Amélie Anota<sup>1,3,4</sup>, Franck Bonnetain<sup>1,3,4,5</sup>, Jean-Philippe Oster<sup>6</sup>, Eric Pichon<sup>7</sup>, Marie Wislez<sup>8</sup>, Jérôme Dauba<sup>9</sup>, Didier Debieuvre<sup>10</sup>, Pierre-Jean Souquet<sup>11</sup>, Laurence Bigay-Game<sup>12</sup>, Olivier Molinier<sup>13</sup>, Eric Dansin<sup>14</sup>, Michel Poudenx<sup>15</sup>, Bernard Milleron<sup>16</sup>, Franck Morin<sup>16</sup>, Gérard Zalcman<sup>16,17</sup>, Elisabeth Quoix<sup>16,18</sup> and Virginie Westeel<sup>16,19</sup>

**Affiliations:** <sup>1</sup>Methodology and Quality of Life in Oncology Unit, University Hospital of Besançon, Besançon, France. <sup>2</sup>Dept of Medical Oncology, University Hospital of Besançon, Besançon, France. <sup>3</sup>EA 3181 University of Franche-Comté, Besançon, France. <sup>4</sup>Platform Quality of Life and Cancer, France. <sup>5</sup>EORTC QOL Group, Brussels, Belgium. <sup>6</sup>Chest Disease Dept, Hospital of Colmar, Colmar, France. <sup>7</sup>Chest Disease Dept, University Hospital of Tours, Tours, France. <sup>8</sup>Chest Disease Dept, Assistance Publique Hôpitaux de Paris (AP-HP), Tenon Hospital, Paris, France. <sup>9</sup>Oncology Dept, University Hospital of Mont-de-Marsan, Mont-de-Marsan, France. <sup>10</sup>Chest Disease Dept, Hospital of Mulhouse, Mulhouse, France. <sup>11</sup>Chest Disease Dept, University Hospital of Lyon-Sud, Lyon, France. <sup>12</sup>Chest Disease Dept, University Hospital of Toulouse, Toulouse, France. <sup>13</sup>Chest Disease Dept, Hospital of Le Mans, Le Mans, France. <sup>14</sup>Cancer Centre O. Lambret, Lille, France. <sup>15</sup>Cancer Centre A. Lacassagne, Nice, France. <sup>16</sup>Intergroupe Francophone de Cancérologie Thoracique (IFCT), Paris, France. <sup>17</sup>Dept of Thoracic Oncology, Assistance, Publique Hôpitaux de Paris, Bichat-Claude Bernard Hospital, Paris, France. <sup>18</sup>Chest Disease Department, University Hospital of Strasbourg, Strasbourg, France. <sup>19</sup>Chest Disease Dept, University Hospital of Besançon, Besançon, France.

**Correspondence:** Frédéric Fiteni, University Hospital of Besançon, Department of Medical Oncology, 3 Boulevard Fleming, 25000, Besançon, France. E-mail: fredericfiteni@gmail.com

**ABSTRACT** In the *Intergroupe Francophone de Cancérologie Thoracique* 0501 trial the carboplatin-paclitaxel chemotherapy increased toxicity (most frequent, decreased neutrophil count, asthenia). We longitudinally compared health-related quality of life (HRQoL) of the two treatment arms.

In total, 451 patients aged 70–89 years with advanced non-small cell lung cancer (NSCLC) were randomly assigned to receive carboplatin plus paclitaxel or vinorelbine or gemcitabine. HRQoL was assessed by means of the European Organisation for Research and Treatment of Cancer QLQ-C30 questionnaire at baseline, week 6 and week 18.

Using a five-point decrease as the minimal clinically important difference, patients treated with the chemotherapy doublet exhibited a significant longer time until definitive deterioration (TUDD) for two HRQoL dimensions: physical functioning (median TUDD: 2.04 for the doublet *versus* 1.71 months for monotherapy; log-rank  $p=0.01$ ) and nausea and vomiting (median: not reached *versus* 4.83, respectively; log-rank  $p=0.046$ ). Cox multivariate analysis revealed the carboplatin and paclitaxel arm to be independently associated with longer TUDD for these two HRQoL dimensions. In addition, TUDD didn't significantly differ between the two arms for all the other HRQoL dimensions.

The chemotherapy doublet did not reduce TUDD in elderly patients with advanced NSCLC. Moreover, TUDD was prolonged for two HRQoL dimensions, namely physical functioning and nausea and vomiting.



@ERSpublications

**Carboplatin plus paclitaxel chemotherapy does not reduce quality of life in elderly patients with advanced NSCLC** <http://ow.ly/obcW300xe59>

This article has supplementary material available from [erj.ersjournals.com](http://erj.ersjournals.com)

Received: Oct 13 2015 | Accepted after revision: April 26 2016

Copyright ©ERS 2016

## Introduction

Increases in life expectancy and the strong association of lung cancer frequency with ageing have led to a significant rise of this cancer type in the elderly. In a community hospital-based survey in France, patients aged 70 years and over, presenting with pathological confirmed lung cancer, represented 32% of the total 5667 patients recorded during year 2000, and those aged over 80 years represented 18.1% [1, 2].

Between 2006 and 2009, the *Intergroupe Francophone de Cancérologie Thoracique* (IFCT; French Intergroup of Thoracic Oncology) conducted a phase III trial (IFCT-0501 trial) involving elderly patients with advanced non-small cell lung cancer (NSCLC) in order to compare the carboplatin and weekly paclitaxel doublet regimen to single-agent chemotherapy with gemcitabine or vinorelbine [3]. Median overall survival was 10.3 months for the doublet and 6.2 months for the monotherapy (hazard ratio 0.64, 95% CI 0.52–0.78;  $p < 0.0001$ ). Several grade 3–4 adverse events were, however, found to be significantly increased in the doublet arm, namely neutropenia, febrile neutropenia, anaemia, thrombocytopenia, asthenia and sensory neuropathy. Since this study was published, the European Society for Medical Oncology (ESMO) [4] and the National Comprehensive Cancer Network (NCCN) [5] have recommended platinum-based chemotherapy as the preferred option for patients aged 70 years or over with a performance status of 0–1, as well as selected performance status 2 patients with adequate organ function. A single-agent approach may still be the recommended treatment for unfit or comorbid patients who are more likely to suffer from significantly more treatment-related adverse events [6].

Health-related quality of life (HRQoL) reflects the patient-perceived evaluation of one's health, including physical, emotional and social dimensions as well as symptoms due to disease or treatment. The Food and Drug Administration considers HRQoL to be a significant endpoint for assessing direct clinical benefits for the patients [7, 8]. Some publications have focused on the comparison between patient and clinician reporting of the treatment-related adverse events experienced by cancer patients, which have demonstrated underreporting of adverse events by clinicians [9–12].

Therefore, given the slight increase in adverse event frequency observed in the IFCT-0501 trial, we sought to longitudinally compare HRQoL scores between both treatment arms of the IFCT-0501 trial [13].

## Methods

### *Patients and study design*

Patients were eligible if they were aged 70–89 years and presented with stage IV NSCLC or a stage III disease unsuitable for radical radiation therapy and a performance status  $\leq 2$ . Patients were randomly assigned 1:1 (minimisation) to either four 4-week cycles of carboplatin (day 1) plus paclitaxel (days 1, 8 and 15) or five 3-week cycles of vinorelbine or gemcitabine (days 1 and 8). The protocol was approved by the *Comité de Protection des Personnes* (a French research ethics board) of Ile-de-France Aulnay-sous-Bois, France. The trial was authorised by the French National Authority for Health. All enrolled patients provided written informed consent. The study design has been described in detail in a previous publication [3].

### *Health-related quality of life*

HRQoL was a secondary endpoint of the trial, assessed using the European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire (QLQ)-C30 [14] at randomisation and then again at 6 and 18 weeks. Additional questionnaires were given to 65 (28.6%) patients in the monotherapy arm and 88 (39.2%) in the doublet chemotherapy arm at 12 weeks after randomisation, the results of which were included in the analysis. The questionnaires were filled in by the patients themselves at the hospital on paper prior to a medical visit.

The QLQ-C30 is a cancer-specific tool composed of 30 items [15–17], allowing to assess five functional scores (physical, role, cognitive, social and emotional), a global quality-of life-score and nine symptom scores (nausea and vomiting, pain, fatigue, dyspnoea, sleeping disturbances, appetite loss, constipation, diarrhoea and financial difficulties) [14]. These scores were generated according to the EORTC Scoring Manual and thus were standardised on a 0 to 100 scale in order that a high score reflects a high functional level, a high global quality of life level and a high symptomatic level [18]. If at least half of the items of one given dimension were completed, the score could be estimated considering that missing items did not differ from items answered (corresponding to simple imputation by the mean). If more than half of the items were missing, the score could not be computed and thus was missing.

---

Support statement: This work was supported by the Intergroupe Francophone de Cancérologie Thoracique and the Ligue Nationale Contre le Cancer.

Conflict of interest: Disclosures can be found alongside the online version of this article at [erj.ersjournals.com](http://erj.ersjournals.com)

### **Statistical methods**

All analyses were performed according to the modified intention-to-treat (ITT) principle, *i.e.* an ITT population with at least one available baseline HRQoL questionnaire.

To determine the profile of baseline missing HRQoL data, patients who entirely completed the baseline HRQoL questionnaire (*i.e.* with all items completed) were compared with other patients (*i.e.* those with at least one missing item at baseline) according to primary clinical and medical patients characteristics collected at baseline. Qualitative/categorical variables have been described using frequencies and percentages and then being compared by group (*i.e.* according to the availability or not of the baseline HRQoL questionnaire and then according to treatment arm) using the Chi-squared or Fisher's exact test when expected numbers were less than 5%. Continuous/quantitative variables were described using mean $\pm$ SD or median (range) and compared by group using a t-test or the Mann-Whitney non-parametric test if a normal distribution of the variable was not respected.

The minimal clinically important difference (MCID) for HRQoL scores was fixed to 5 points [19, 20]. The time until definitive HRQoL score deterioration (TUDD) of a score was defined as the interval between randomisation and the deterioration  $\geq$ 5 points in the QLQ-C30 score compared to the HRQoL score at baseline, with no further improvement in HRQoL score  $\geq$ 5 points or if a patient dropped out after this decrease, resulting in missing data. Patients were censored at the last follow-up when no deterioration in HRQoL score compared with baseline was observed or in cases where a deterioration was observed but followed by a significant improvement in HRQoL score compared to baseline. Patients with a baseline score but no follow-up score have been included in the analysis, though, were censored immediately after baseline.

The TUDD curves were calculated using the Kaplan-Meier estimation and described using medians and 95% confidence interval. TUDD curves were compared using log-rank tests. The univariate Cox model was applied to calculate the HR with its 95% confidence interval. We assessed the following variables: treatment arm, age, sex, performance status, smoking status, mini-mental state (MMS) examination score, activities of daily living (ADL) questionnaire score, Charlson comorbidity index (CCI) score, body mass index (BMI), disease stage and histology [3]. All variables with a p-value  $<$ 0.1 were included in a multivariate Cox regression model in order to identify factors independently associated with TUDD.

We performed sensitivity analyses with the aim of evaluating the different definitions of TUDD. These analyses were repeated with 10-point differences in scores for the MCID.

The analysis was conducted according to the patient-reported outcomes CONSORT (Consolidated Standards of Reporting Trials) statement [21].

A p-value of 0.05 or lower was considered statistically significant. All analyses were performed using SAS software, Version 9.3 (SAS Institute), and R software (Version 2.10.1).

## **Results**

### **Patients**

Between April 2006 and December 2009, 451 patients were enrolled. Of them, 226 were assigned single-agent chemotherapy (62 with vinorelbine and 164 with gemcitabine) and 225 received doublet chemotherapy [3].

### **HRQoL compliance and scores at baseline**

The total number of patients who completed the entire questionnaire was 361 (80.04%) at baseline, 191 (50.9%) at 6 weeks, 69 (25.6%) at 12 weeks, and 86 (39.1%) at 18 weeks.

No difference was observed in terms of baseline characteristics of the patients between patients who completed the entire baseline HRQoL questionnaire and those who did not, suggesting that baseline missing data could be considered as missing completely at random (table 1). The mean HRQoL scores at baseline were similar between the two treatment arms (table 2).

The total number of available questionnaires (*i.e.* the number of questionnaires with at least one score of HRQoL that could be calculated) was 421 at baseline, 315 (74.8%) at 6 weeks, 153 (36.3%) at 12 weeks, and 250 (53.4%) at 18 weeks (figure 1).

### **TUDD with a MCID $>$ 5 points**

The results in table 3 showed that the TUDD with an MCID  $\geq$ 5 points in HRQoL score was significantly longer under the chemotherapy doublet compared with monotherapy for two HRQoL dimensions: physical functioning and nausea and vomiting.



TABLE 1 Patient characteristics according to the completion of the baseline EORTC QLQ-C30

	Questionnaire complete at baseline	Questionnaire not complete at baseline	Fisher exact test p-value
<b>Patients n</b>	361	90	
<b>Age years</b>			
<77	180 (49.9)	48 (53.3)	0.56
≥77	181 (50.1)	42 (46.7)	
<b>Sex</b>			
Male	270 (74.8)	63 (70.0)	0.35
Female	91 (25.2)	27 (30.0)	
<b>Performance status</b>			
0–1	267 (74.0)	63 (70.0)	0.15
2	93 (25.7)	27 (30.0)	
Unknown	1 (0.3)	0 (0)	
<b>Smoking status</b>			
Never smoked	73 (20.2)	21 (23.3)	0.51
Ever smoked	288 (80.0)	69 (76.7)	
<b>MMS</b>			
≤20	29 (8.0)	6 (6.7)	0.82
>20	330 (91.4)	76 (84.4)	
Unknown	2 (0.6)	8 (8.9)	
<b>ADL</b>			
<6	66 (18.3)	22 (24.4)	0.12
6	288 (79.8)	62 (68.9)	
Unknown	7 (1.9)	6 (6.7)	
<b>CCI</b>			
≤2	268 (74.2)	73 (81.1)	0.17
>2	93 (25.8)	17 (18.9)	
<b>BMI</b>			
≤20	43 (11.9)	9 (10.0)	0.77
20<BMI≤30	276 (76.5)	72 (80.0)	
>30	42 (11.6)	9 (10.0)	
<b>Stage</b>			
III A–III B	70 (19.4)	17 (18.9)	0.90
IV	291 (80.6)	73 (81.1)	
<b>Histology</b>			
Adenocarcinoma	184 (51.0)	45 (50.0)	0.63
Squamous	123 (34.0)	28 (31.1)	
Other	54 (15.0)	17 (18.9)	

Data are presented as n (%), unless otherwise stated. MMS: mini-mental state; ADL: activities of daily life; CCI: Charlson's comorbidity index; BMI: body mass index.

For the physical functioning score, 72 patients in the doublet arm and 91 in the single-agent arm experienced a definitive deterioration ≥5 points. The median TUDD was 2.04 months (95% CI 1.87–3.88) and 1.71 months (95% CI 1.58–1.91), respectively (HR 0.67, 95% CI 0.49–0.91; p=0.01) (figure 2a). Cox multivariate analysis revealed that the carboplatin and paclitaxel doublet was independently associated with longer TUDD for the physical functioning dimension (HR 0.57, 95% CI 0.42–0.78; p=0.0079) (table 4).

For the nausea and vomiting score, 10 and 19 patients experienced a definitive deterioration ≥5 points in each arm, respectively. Median TUDD was not reached (NR) in the doublet arm and was 4.83 months (95% CI 4.70–NR) in the monotherapy arm (HR 0.46, 95% CI 0.21–1.00; p=0.046) (figure 2b). Cox multivariate analysis revealed that the carboplatin and paclitaxel doublet was independently associated with longer TUDD for the nausea and vomiting dimension (HR 0.39, 95% CI 0.18–0.87; p=0.02) (table 5).

For two other dimensions, namely fatigue (FA) and social functioning (SF), TUDD was longer for the chemotherapy doublet arm, with borderline significance. For the fatigue score, 33 patients in the doublet chemotherapy group and 44 in the monotherapy group experienced a definitive deterioration ≥5 points. The median TUDD was 5.03 months (95% CI 4.44–NR) in the chemotherapy doublet arm and 4.37 months (95% CI 4.24–4.80) in the monotherapy arm (HR 0.65, 95% CI 0.41–1.03; p=0.06). Cox multivariate analysis demonstrated a trend towards longer TUDD for the fatigue dimension with the carboplatin and paclitaxel doublet (HR 0.67, 95% CI 0.42–1.05; p=0.08) (supplementary table S1).

TABLE 2 Health-related quality of life scores at baseline by treatment arm

QLQ-C30 scores	Monotherapy group (n=226)			Doublet chemotherapy group (n=225)			p-value
	n	Mean±sd	Median (range)	n	Mean±sd	Median (range)	
Global quality of life	214	55.4±18.4	58.3 (8.3–100)	206	58.2±19	58.3 (16.7–100)	0.12
Physical functioning	214	69.1±22.5	73.3 (6.7–100)	206	69.3±22.5	73.3 (0–100)	0.95
Role functioning	214	63±35.7	66.7 (0–100)	206	69.1±32.3	75 (0–100)	0.068
Emotional functioning	213	71.3±21.3	75 (0–100)	207	73.1±23	75 (0–100)	0.42
Cognitive functioning	214	81.2±21.9	83.3 (0–100)	207	83.7±20.2	83.3 (0–100)	0.2
Social functioning	209	75.6±31.8	100 (0–100)	202	80.4±29	100 (0–100)	0.11
Fatigue	213	45.5±28	33.3 (0–100)	207	41.4±27.6	33.3 (0–100)	0.14
Nausea and vomiting	214	5.4±14.5	0 (0–100)	207	5.6±14.6	0 (0–100)	0.86
Pain	214	29±29.8	16.7 (0–100)	206	26.7±29.9	16.7 (0–100)	0.44
Dyspnoea	213	47.4±34.1	33.3 (0–100)	206	40.8±35.1	33.3 (0–100)	0.050
Sleeping disturbances	214	29.8±32	33.3 (0–100)	206	28±32.8	33.3 (0–100)	0.58
Appetite loss	213	34.7±29.7	33.3 (0–100)	206	36.9±31.8	33.3 (0–100)	0.55
Constipation	214	24.9±32.2	0 (0–100)	206	26.2±31.8	0 (0–100)	0.68
Diarrhoea	211	7.3±18.1	0 (0–100)	206	7.8±18.7	0 (0–100)	0.78
Financial difficulties	212	5.2±16.5	0 (0–100)	204	3.6±13.6	0 (0–100)	0.28

For the SF dimension, 55 patients in the doublet chemotherapy group and 65 in the monotherapy group experienced a definitive deterioration  $\geq 5$  points. The median TUDD was 4.21 months (95% CI 2.33–4.37) in the chemotherapy doublet arm and 1.91 months (95% CI 1.74–4.01) in the monotherapy arm (HR 0.73, 95% CI 0.51–1.05;  $p=0.049$ ). Cox multivariate analysis revealed that the carboplatin and paclitaxel doublet

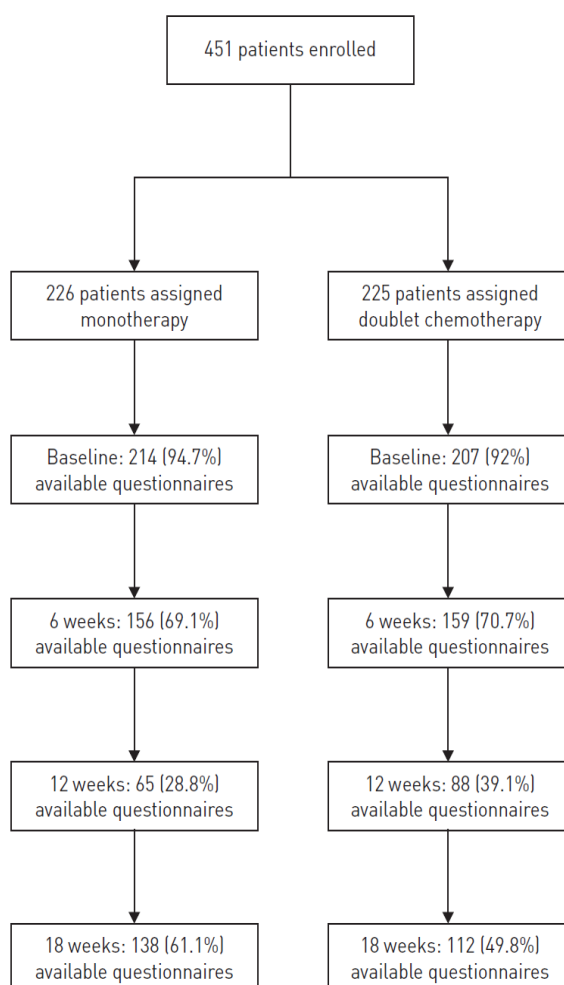


FIGURE 1 Number of available questionnaires (i.e. the number of questionnaires with at least one score of health related quality of life that can be calculated).

TABLE 3 Time until definitive deterioration with minimal clinically important difference  $\geq 5$  points and 10 points according to treatment arm

	MCID $\geq 5$ points			MCID $\geq 10$ points		
	Events	Median (95% CI) TUDD months	Log-rank p-value	Events	Median (95% CI) TUDD months	Log-rank p-value
<b>Global quality of life</b>						
Monotherapy	73	1.91 (1.71–4.01)	0.35	51	4.24 (3.98–NR)	0.84
Doublet chemotherapy	69	2.33 (1.87–4.21)		52	4.27 (2.33–NR)	
<b>Physical functioning</b>						
Monotherapy	91	1.71 (1.58–1.91)	<b>0.010</b>	70	2.14 (1.74–4.17)	0.23
Doublet chemotherapy	72	2.04 (1.87–3.88)		64	3.81 (2.04–4.21)	
<b>Role functioning</b>						
Monotherapy	59	3.98 (1.91–4.50)	0.94	59	3.98 (1.91–4.50)	0.94
Doublet chemotherapy	66	3.68 (1.94–4.14)		66	3.68 (1.94–4.14)	
<b>Emotional functioning</b>						
Monotherapy	49	4.24 (3.19–NR)	0.32	40	NR (3.88–NR)	0.27
Doublet chemotherapy	45	4.40 (4.04–NR)		35	4.53 (4.37–NR)	
<b>Cognitive functioning</b>						
Monotherapy	55	4.24 (1.91–4.67)	0.75	55	4.24 (1.91–4.70)	0.75
Doublet chemotherapy	58	4.14 (3.71–4.53)		58	4.14 (3.71–4.53)	
<b>Social functioning</b>						
Monotherapy	65	1.91 (1.74–4.01)	<b>0.09</b>	65	1.91 (1.74–4.01)	0.09
Doublet chemotherapy	55	4.21 (2.33–4.37)		55	4.21 (2.33–4.37)	
<b>Fatigue</b>						
Monotherapy	44	4.37 (4.24–4.80)	0.06	44	4.37 (4.24–4.80)	<b>0.05</b>
Doublet chemotherapy	33	5.03 (4.44–NR)		32	5.03 (4.44–NR)	
<b>Nausea and vomiting</b>						
Monotherapy	19	4.83 (4.70–NR)	<b>0.046</b>	19	4.83 (4.70–NR)	<b>0.046</b>
Doublet chemotherapy	10	NR (5.03–NR)		10	NR (5.03–NR)	
<b>Pain</b>						
Monotherapy	50	4.30 (4.01–4.83)	0.16	50	4.30 (4.01–4.83)	0.10
Doublet chemotherapy	40	NR (4.24–NR)		40	NR (4.23–NR)	
<b>Dyspnoea</b>						
Monotherapy	42	4.50 (4.14–4.86)	0.40	42	4.50 (4.14–4.86)	0.40
Doublet chemotherapy	51	4.07 (2.27–5.03)		51	4.07 (2.27–5.03)	
<b>Sleeping disturbances</b>						
Monotherapy	41	4.83 (4.14–4.83)	0.27	41	4.83 (4.14–4.83)	0.27
Doublet chemotherapy	33	NR (4.30–NR)		33	NR (4.31–NR)	
<b>Appetite loss</b>						
Monotherapy	36	4.66 (4.24–4.83)	0.93	36	4.67 (4.24–4.83)	0.93
Doublet chemotherapy	40	4.86 (4.01–NR)		40	4.86 (4.01–NR)	
<b>Constipation</b>						
Monotherapy	28	4.67 (4.37–NR)	0.79	28	4.67 (4.37–NR)	0.79
Doublet chemotherapy	35	5.03 (4.37–NR)		35	5.03 (4.37–NR)	
<b>Diarrhoea</b>						
Monotherapy	16	4.83 (4.67–NR)	0.25	16	4.83 (4.67–NR)	0.25
Doublet chemotherapy	12	NR (NR–NR)		12	NR (NR–NR)	
<b>Financial difficulties</b>						
Monotherapy	10	NR (4.83–NR)	0.31	10	NR (4.83–NR)	0.32
Doublet chemotherapy	6	NR (NR–NR)		6	NR (NR–NR)	

NR: not reached.

prolonged TUDD for the SF dimension, although only with borderline statistical significance (HR 0.71, 95% CI 0.49–1.02;  $p=0.06$ ) (supplementary table S2).

For the other HRQoL dimensions, the chemotherapy doublet arm did not alter TUDD, compared with monotherapy.

#### Sensitivity analyses

Table 3 shows the results with a MCID  $\geq 10$  points in HRQoL score.

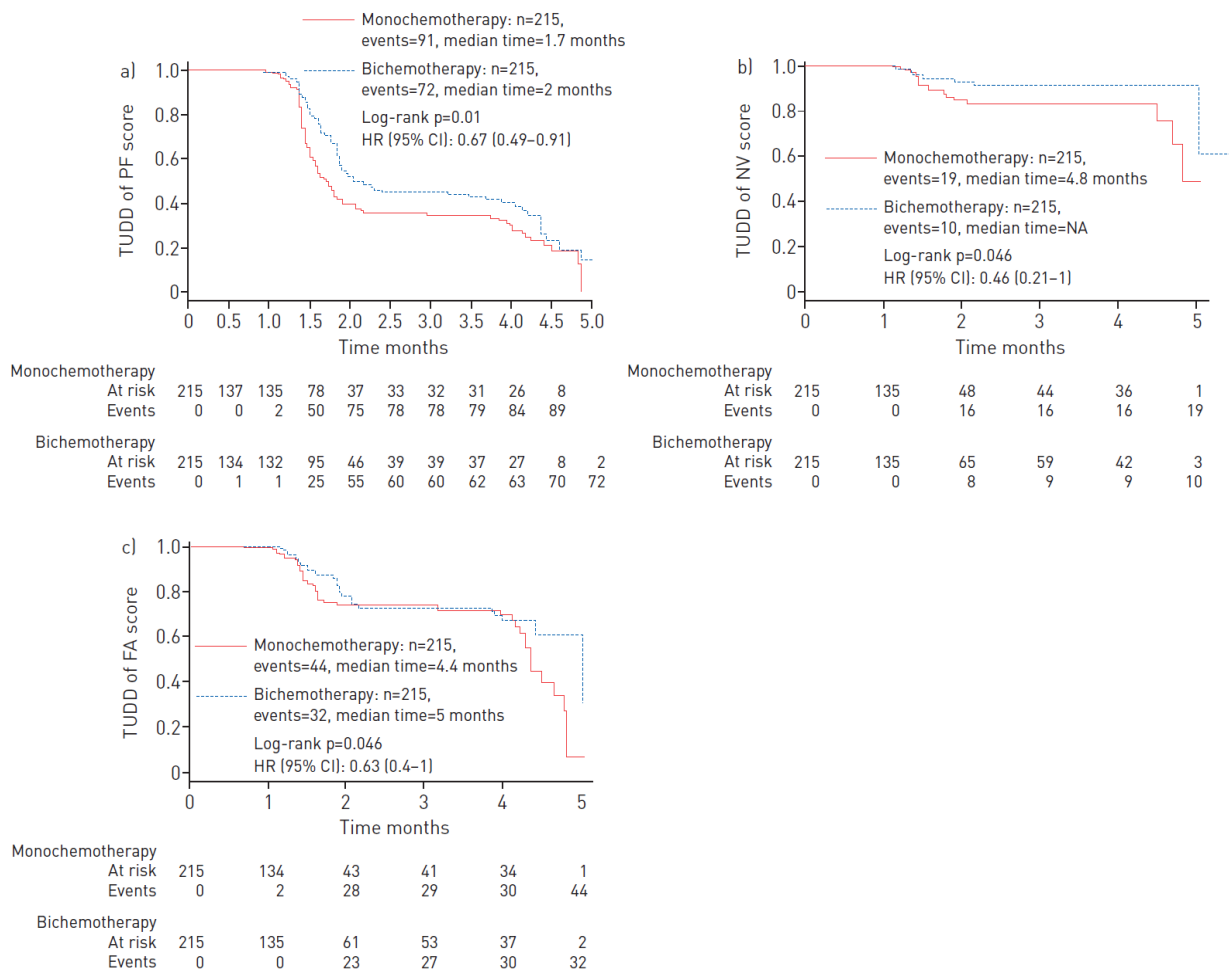


FIGURE 2 a) Time until definitive deterioration (TUDD) of physical functioning (PF) score according to treatment arm (Kaplan-Meier estimation) with minimal clinically important difference (MCID)  $\geq 5$ . b) TUDD of nausea and vomiting (NV) score according to treatment arm (Kaplan-Meier estimation) with MCID  $\geq 5$  points. c) TUDD of fatigue (FA) score according to treatment arm (Kaplan-Meier estimation) with MCID  $\geq 10$  points.

TUDD was significantly longer under the doublet when compared with single-agent chemotherapy for the FA dimension. For the FA score, 32 patients in the doublet chemotherapy arm and 44 in the monotherapy arm experienced a definitive deterioration  $\geq 10$  points. The median TUDD was 5.03 months (95% CI 4.44 months–NR) in the doublet chemotherapy arm and 4.37 months (95% CI 4.24–4.80 months) in the monotherapy arm (HR 0.63; 95% CI 0.40–0.99;  $p=0.046$ ) (figure 2c). Cox multivariate analysis once more demonstrated that the carboplatin and paclitaxel doublet was independently associated with longer TUDD for the FA dimension (HR 0.60; 95% CI 0.49–0.74;  $p<0.001$ ) (table 6).

Similar results to the primary analysis were observed for the other dimensions.

### Discussion

The IFCT-0501 trial demonstrated that, when compared to a monotherapy regimen, the carboplatin/paclitaxel therapy combination improved overall survival in elderly patients with advanced NSCLC. However, the safety of carboplatin/paclitaxel was slightly less favourable than that of monotherapy. Studies have revealed that most oncologists and patients are unwilling to prolong survival at the expense of worsening HRQoL [22, 23]. It was therefore important to assess the HRQoL results as a secondary endpoint in the IFCT-0501 trial. Our study highlighted that TUDD values with an MCID  $\geq 5$  points in HRQoL scores were significantly longer with doublet chemotherapy than with monotherapy for two HRQoL dimensions, namely physical functioning and nausea and vomiting. Moreover, for two other dimensions, SF and FA, TUDD was favoured by the chemotherapy doublet, though with borderline

TABLE 4 Cox analysis of time until definitive deterioration for physical functioning dimension with minimal clinically important difference  $\geq 5$

	Patients n	Events n	Univariate analysis (n=430)		Multivariate analysis (n=430)	
			HR (95% CI)	p-value	HR (95% CI)	p-value
<b>Total</b>		163				
<b>Sex</b>						
Male	317	110	1.00		1.00	
Female	113	53	1.37 (0.98–1.90)	<b>0.065</b>	1.53 (1.09–2.15)	<b>0.013</b>
<b>Age years</b>						
<77	218	84	1.00			
$\geq 77$	212	79	0.83 (0.61–1.13)	0.24		
<b>Treatment</b>						
Monotherapy	215	91	1.00		1.00	
Doublet chemotherapy	215	72	0.67 (0.49–0.91)	<b>0.01</b>	0.57 (0.42–0.78)	<b>0.008</b>
<b>Performance status score</b>						
0–1	312	127	1.00			
2	117	36	1.25 (0.86–1.81)	0.24		
<b>Smoking status</b>						
Never smoker	91	45	1.00			
Ever smoker	339	118	0.77 (0.55–1.10)	0.15		
<b>Disease stage</b>						
III	82	31	1.00			
IV	348	132	1.09 (0.74–1.62)	0.65		
<b>Histology</b>						
Adenocarcinoma	218	87	1.00			
Squamous	146	48	1.04 (0.73–1.49)			
Other	66	28	0.99 (0.65–1.53)	0.97		
<b>MMS</b>						
$\leq 20$	33	9	1.00			
>20	393	152	0.99 (0.50–1.94)	0.96		
<b>ADL</b>						
<6	83	27	1.00			
6	339	132	0.75 (0.49–1.13)	0.17		
<b>CCI</b>						
$\leq 2$	321	116	1.00		1.00	
>2	109	47	1.63 (1.16–2.29)	<b>0.005</b>	1.50 (1.08–1.12)	<b>0.002</b>
<b>BMI</b>						
$\leq 20$	48	15	1.00			
20<BMI $\leq 30$	332	127	0.88 (0.51–1.50)			
>30	50	21	0.72 (0.37–1.40)	0.59		

MMS: mini-mental state examination questionnaire; ADL: activities of daily living questionnaire; CCI: Charlson's comorbidity index; BMI: body mass index.

significance. Finally, no difference in TUDD was observed for the other HRQoL dimensions when comparing the two treatment arms. For the FA dimension, TUDD with an MCID  $\geq 5$  points was not prolonged in the doublet arm compared with monotherapy, while the TUDD with an MCID  $\geq 10$  points was significantly longer. This could be accounted for by the observation that one patient in the doublet chemotherapy arm exhibited a deterioration of between 6 and 10 points, leading to 32 events using the MCID  $\geq 10$  points and 33 using an MCID  $\geq 5$  points. For the physical functioning dimension, TUDD with an MCID  $\geq 10$  points was no longer significantly increased under the doublet chemotherapy. This could be explained by the fact that less patients exhibited deterioration with an MCID  $\geq 10$  points (64 and 70 for the doublet chemotherapy and monotherapy arm, respectively) compared with an MCID  $\geq 5$  points (72 and 91 for the doublet chemotherapy and monotherapy arm, respectively).

These results suggest that, despite the moderately increased toxicity induced by the chemotherapy doublet, TUDD was, in fact, longer in patients treated with the chemotherapy doublet for the four following HRQoL dimensions: physical functioning, which assesses self-care, mobility and physical activity; SF, which measures how patients can perform their usual work and housework activities; FA; and nausea and vomiting.

Interestingly, we observed significantly more Grade 3 and 4 asthenia with the doublet, while the FA dimension was independently associated with longer TUDD for the carboplatin and paclitaxel arm compared with that

TABLE 5 Cox analysis of time until definitive deterioration for nausea and vomiting dimension with minimal clinically important difference  $\geq 5$  points and with minimal clinically important difference  $\geq 10$  points

	Patients n	Events n	Univariate analysis (n=430)		Multivariate analysis (n=430)	
			HR (95% CI)	p-value	HR (95% CI)	p-value
<b>Total</b>		29				
<b>Sex</b>						
Male	317	15	1.00		1.00	
Female	113	14	2.42 (1.17–5.02)	<b>0.017</b>	2.06 (0.67–6.32)	0.21
<b>Age years</b>						
<77	218	15	1.00			
$\geq 77$	212	14	0.98 (0.47–2.04)	0.96		
<b>Treatment</b>						
Monotherapy	215	19	1.00		1.00	
Doublet chemotherapy	215	10	0.46 (0.21–1.00)	<b>0.05</b>	0.39 (0.18–0.87)	<b>0.02</b>
<b>Performance status score</b>						
0–1	312	21	1.00			
2	117	8	1.44 (0.62–3.32)	0.38		
<b>Smoking status</b>						
Never smoker	91	12	1.00		1.00	
Ever smoker	339	17	0.47 (0.23–0.99)	<b>0.048</b>	1.00 (0.33–3.03)	0.99
<b>Disease stage</b>						
III	82	3	1.00			
IV	348	26	1.96 (0.59–6.51)	0.27		
<b>Histology</b>						
Adenocarcinoma	218	18	1.00			
Squamous	146	6	0.60 (0.24–1.52)			
Other	66	5	0.86 (0.32–2.32)	0.56		
<b>MMS</b>						
$\leq 20$	33	3	1.00			
>20	393	27	0.75 (0.18–3.17)	0.70		
<b>ADL</b>						
<6	83	10	1.00		1.00	
6	339	19	0.31 (0.14–0.67)	<b>0.003</b>	0.36 (0.16–0.83)	<b>0.02</b>
<b>CCI</b>						
$\leq 2$	321	23	1.00			
>2	109	6	0.88 (0.36–2.18)	0.78		
<b>BMI</b>						
$\leq 20$	48	4	1.00			
20 < BMI $\leq 30$	332	20	0.49 (0.17–1.45)			
>30	50	5	0.82 (0.22–3.08)	0.32		

MMS: mini-mental state; ADL: activities of daily life; CCI: Charlson's comorbidity index; BMI: body mass index.

achieved with monotherapy. There are several possible explanations for this apparent contradiction. Firstly, the toxic effects were assessed by the incidence of Grade 3 and 4 toxicities, according to the Common Terminology Criteria for Adverse Events (CTCAE) of the National Cancer Institute (version 3.0), while the TUDD of an HRQoL dimension used a longitudinal approach with a 5- or 10-point MCID. Secondly, the adverse events were evaluated by the physician rather than being self-reported [24–26]. While Grade 3 and 4 nausea and vomiting were reported in 2.7% and 0.9% of the patients in the doublet chemotherapy and monotherapy arms, respectively, TUDD was longer for the carboplatin and paclitaxel arm regarding the nausea and vomiting dimension, compared with the monotherapy arm. Besides the above potential explanations, the nausea and vomiting could also be attributed either directly to the cancer or to other treatments administered for managing cancer-related pain, such as morphine analogues.

One other trial, the Multicenter Italian Lung Cancer in the Elderly Study (MILES) study, analysed HRQoL in elderly NSCLC patients treated with a chemotherapy doublet (combination of vinorelbine plus gemcitabine) compared with single agents (vinorelbine or gemcitabine alone). The authors observed no statistically significant differences in terms of HRQoL between the patients assigned to combination treatment and those assigned to single-drug treatments [27]. Nevertheless, the statistical method used was a simple change-from-baseline approach. The West Japan Thoracic Oncology Group 9904 study compared weekly docetaxel and cisplatin administration *versus* docetaxel monotherapy in elderly patients with advanced NSCLC. This study failed to demonstrate any survival advantage of weekly docetaxel plus

TABLE 6 Cox analysis of time until definitive deterioration for fatigue dimension with minimal clinically important difference  $\geq 10$  points

	Patients n	Events n	Univariate analysis (n=430)		Multivariate analysis (n=430)	
			HR (95% CI)	p-value	HR (95% CI)	p-value
<b>Total</b>		76				
<b>Sex</b>						
Male	317	50	1.00			
Female	113	26	1.28 (0.80–2.06)	0.31		
<b>Age years</b>						
<77	218	38	1.00			
$\geq 77$	212	38	0.95 (0.60–1.48)	0.81		
<b>Treatment</b>						
Monotherapy	215	44	1.00		1.00	
Doublet chemotherapy	215	32	0.63 (0.40–0.99)	<b>0.046</b>	0.60 (0.49–0.74)	<b>&lt; 0.001</b>
<b>Performance status score</b>						
0–1	312	60	1.00			
2	117	16	1.08 (0.62–1.88)	0.80		
<b>Smoking status</b>						
Never smoker	91	24	1.00			
Ever smoker	339	52	0.81 (0.50–1.32)	0.40		
<b>Disease stage</b>						
III	82	12	1.00			
IV	348	64	1.32 (0.71–2.45)	0.38		
<b>Histology</b>						
Adenocarcinoma	218	41	1.00			
Squamous	146	22	1.05 (0.63–1.77)			
Other	66	13	1.07 (0.57–1.99)	0.97		
<b>MMS</b>						
$\leq 20$	33	5	1.00			
>0	393	71	0.79 (0.32–1.97)	0.61		
<b>ADL</b>						
<6	83	13	1.00			
6	339	63	0.90 (0.49–1.64)	0.73		
<b>CCI</b>						
$\leq 2$	321	67	1.00		1.00	
>2	109	9	0.49 (0.24–0.98)	<b>0.044</b>	1.09 (0.86–1.38)	0.50
<b>BMI</b>						
$\leq 20$	48	11	1.00			
20 < BMI $\leq 30$	332	56	0.51 (0.27–0.98)			
>30	50	9	0.48 (0.20–1.17)	0.11		

MMS: mini-mental state; ADL: activities of daily life; CCI: Charlson's comorbidity index; BMI: body mass index.

cisplatin over docetaxel monotherapy. Moreover, HRQoL was not favoured by the doublet chemotherapy [28]. The chemotherapy delivery was inferior in patients receiving the cisplatin–docetaxel combination compared with those treated with single-agent docetaxel, suggesting that a carboplatin-based combination should be recommended in elderly patients.

Time-to-HRQoL score deterioration approaches tend to be extensively used in oncology phase III clinical trials [29–33]. These models have the advantage of producing clinically meaningful results for the clinicians and being less affected by missing data than a classical mixed model of analysis of variance for repeated measures [13]. One limitation of our study consisted of the HRQoL analysis being only a secondary endpoint, with no *a priori* statistical hypothesis allowing for power calculation. Then, the treatment effect on HRQoL was small especially for FA and physical functioning dimensions. This small treatment effect may be explained in part by a lack of statistical power because assessment of HRQoL was conducted four times and there were missing data. Nevertheless, there was a difference of 69 days in the median values of TUDD in the social functioning dimension (borderline significant) which may be clinically interesting for the patient.

Moreover, our study shows that the chemotherapy doublet arm did not alter TUDD, compared with monotherapy. Therefore, these findings provide added support for the benefit of doublet chemotherapy in this palliative patient population which can help clinicians in decision making.

In conclusion, this study showed that, compared with vinorelbine or gemcitabine monotherapy, carboplatin–paclitaxel for elderly patients with advanced NSCLC was associated with a statistically significant longer TUDD for several HRQoL dimensions, while similar TUDD were observed between both treatments for all other HRQoL dimensions. These findings add support for claims of the benefit of the carboplatin weekly paclitaxel doublet chemotherapy in this elderly patient population, proving a significant increase in overall survival. Although at the cost of a slight increase in toxicity, this regimen was proven to induce significant improvement in important HRQoL dimensions and caused no alterations in the other dimensions.

## References

- 1 Piquet J, Blanchon F, Grivaux M, *et al.* Le cancer bronchique primitif du sujet âgé en France. [Primary bronchial carcinoma in elderly subjects in France]. *Rev Mal Respir* 2003; 20: 691–699.
- 2 Quoix E, Westeel V, Zalcman G, *et al.* Chemotherapy in elderly patients with advanced non-small cell lung cancer. *Lung Cancer Amst Neth* 2011; 74: 364–368.
- 3 Quoix E, Zalcman G, Oster J-P, *et al.* Carboplatin and weekly paclitaxel doublet chemotherapy compared with monotherapy in elderly patients with advanced non-small-cell lung cancer: IFCT-0501 randomised, phase 3 trial. *Lancet* 2011; 378: 1079–1088.
- 4 Peters S, Adjei AA, Gridelli C, *et al.* Metastatic non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2012; 23: Suppl 7, vii56–vii64.
- 5 Ganti AK, deShazo M, Weir AB III, *et al.* Treatment of non-small cell lung cancer in the older patient. *J Natl Compr Cancer Netw* 2012; 10: 230–239.
- 6 Reck M, Papat S, Reinmuth N, *et al.* Metastatic non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Ann Oncol* 2014; 25: Suppl 3, iii27–iii39.
- 7 Beitz J, Gnecco C, Justice R. Quality-of-life end points in cancer clinical trials: the U.S. Food and Drug Administration perspective. *J Natl Cancer Inst Monogr* 1996; 20: 7–9.
- 8 Johnson JR, Temple R. Food and Drug Administration requirements for approval of new anticancer drugs. *Cancer Treat Rep* 1985; 69: 1155–1159.
- 9 Fromme EK, Eilers KM, Mori M, *et al.* How accurate is clinician reporting of chemotherapy adverse effects? A comparison with patient-reported symptoms from the Quality-of-Life Questionnaire C30. *J Clin Oncol* 2004; 22: 3485–3490.
- 10 Basch E, Iasonos A, McDonough T, *et al.* Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol* 2006; 7: 903–909.
- 11 Cirillo M, Venturini M, Ciccarelli L, *et al.* Clinician versus nurse symptom reporting using the National Cancer Institute-Common Terminology Criteria for Adverse Events during chemotherapy: results of a comparison based on patient's self-reported questionnaire. *Ann Oncol* 2009; 20: 1929–1935.
- 12 Di Maio M, Basch E, Bryce J, *et al.* Patient-reported outcomes in the evaluation of toxicity of anticancer treatments. *Nat Rev Clin Oncol* 2016; 13: 319–325.
- 13 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 Int J Qual Life Asp Treat Care Rehabil* 2015; 24: 5–18.
- 14 Aaronson NK, Ahmedzai S, Bergman B, *et al.* The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* 1993; 85: 365–376.
- 15 Ringdal GI, Ringdal K. Testing the EORTC Quality of Life Questionnaire on cancer patients with heterogeneous diagnoses. *Qual Life Res* 1993; 2: 129–140.
- 16 Anderson RT, Aaronson NK, Wilkin D. Critical review of the international assessments of health-related quality of life. *Qual Life Res* 1993; 2: 369–395.
- 17 Hjermstad MJ, Fossa SD, Bjordal K, *et al.* Test/retest study of the European Organization for Research and Treatment of Cancer Core Quality-of-Life Questionnaire. *J Clin Oncol* 1995; 13: 1249–1254.
- 18 EORTC QLQ-C30 Scoring Manual. Available from [www.eortc.be/qol/files/SCManualQLQ-C30.pdf](http://www.eortc.be/qol/files/SCManualQLQ-C30.pdf) Date last updated: 2001. Date last accessed: May 31, 2016.
- 19 Osoba D, Rodrigues G, Myles J, *et al.* Interpreting the significance of changes in health-related quality-of-life scores. *J Clin Oncol* 1998; 16: 139–144.
- 20 Maringwa JT, Quinten C, King M, *et al.* Minimal important differences for interpreting health-related quality of life scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials. *Support Care Cancer* 2011; 19: 1753–1760.
- 21 Calvert M, Blazeby J, Altman DG, *et al.* Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA* 2013; 309: 814–822.
- 22 Brundage M, Leis A, Bezzak A, *et al.* Cancer patients' preferences for communicating clinical trial quality of life information: a qualitative study. *Qual Life Res* 2003; 12: 395–404.
- 23 Bezzak A, Ng P, Skeel R, *et al.* Oncologists' use of quality of life information: results of a survey of Eastern Cooperative Oncology Group physicians. *Qual Life Res* 2001; 10: 1–13.
- 24 Sprangers MA, Aaronson NK. The role of health care providers and significant others in evaluating the quality of life of patients with chronic disease: a review. *J Clin Epidemiol* 1992; 45: 743–760.
- 25 Moinpour CM, Lyons B, Schmidt SP, *et al.* Substituting proxy ratings for patient ratings in cancer clinical trials: an analysis based on a Southwest Oncology Group trial in patients with brain metastases. *Qual Life Res* 2000; 9: 219–231.
- 26 Slevin ML, Plant H, Lynch D, *et al.* Who should measure quality of life, the doctor or the patient? *Br J Cancer* 1988; 57: 109–112.
- 27 Gridelli C, Perrone F, Gallo C, *et al.* Chemotherapy for elderly patients with advanced non-small-cell lung cancer: the Multicenter Italian Lung Cancer in the Elderly Study (MILES) phase III randomized trial. *J Natl Cancer Inst* 2003; 95: 362–372.



- 28 Abe T, Takeda K, Ohe Y, *et al.* Randomized phase III trial comparing weekly docetaxel plus cisplatin *versus* docetaxel monotherapy every 3 weeks in elderly patients with advanced non-small-cell lung cancer: The Intergroup Trial JCOG0803/WJOG4307L. *J Clin Oncol* 2015; 33: 575–581.
- 29 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: 2753–2762.
- 30 Hamidou Z, Dabakuyo TS, Mercier M, *et al.* Time to deterioration in quality of life score as a modality of longitudinal analysis in patients with breast cancer. *Oncologist* 2011; 16: 1458–1468.
- 31 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* 2013; 31: 23–29.
- 32 Burris HA III, Lebrun F, Rugo HS, *et al.* Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. *Cancer* 2013; 119: 1908–1915.
- 33 Kabbinavar FF, Wallace JF, Holmgren E, *et al.* Health-related quality of life impact of bevacizumab when combined with irinotecan, 5-fluorouracil, and leucovorin or 5-fluorouracil and leucovorin for metastatic colorectal cancer. *Oncologist* 2008; 13: 1021–1029.

**Supplementary table 1.** Cox analysis of TUDD for fatigue dimension with MCID  $\geq 5$

	number of patients	Number of events	Univariate analysis (n=430)		Multivariate analysis (n=431)	
			HR (95% CI)	p-value	HR (95% CI)	p-value
77						
<b>Sex</b>						
Male	317		1.00			
Female	113		1.26 (0.78-2.02)	0.35		
<b>Age (years)</b>						
< 77	218		1.00			
$\geq 77$	212		0.92 (0.59-1.44)	0.72		
<b>Treatment</b>						
Monotherapy	215		1.00		1.00	
Doublet chemotherapy	215		0.65 (0.41-1.03)	0.06	0.67 (0.42-1.05)	0.08
<b>Performance status score</b>						
0-1	312		1.00			
2	117		1.06 (0.61-1.84)	0.85		
<b>Smoking status</b>						
Never smoker	91		1.00			
Ever smoker	339		0.83 (0.51-1.34)	0.44		
<b>Disease stage</b>						
III	82		1.00			
IV	348		1.33 (0.72-2.48)	0.36		
<b>Histology</b>						
Adenocarcinoma	218		1.00			
Squamous	146		1.10 (0.66-1.84)			
Other	66		1.07 (0.57-1.99)	0.93		
<b>MMS</b>						
$\leq 20$	33		1.00			
> 20	393		0.80 (0.32-2.00)	0.63		
<b>ADL</b>						
< 6	83		1.00			
6	339		0.83 (0.47-1.49)	0.54		
<b>CCI</b>						
$\leq 2$	321		1.00		1.00	
> 2	109		0.48 (0.24-0.97)	0.04	0.49 (0.24-0.98)	0.045
<b>BMI</b>						
$\leq 20$	48		1.00		1.00	
20<BMI $\leq$ 30	332		0.52 (0.27-1.00)		0.59 (0.31-1.14)	
>30	50		0.48 (0.20-1.17)	0.13	0.55 (0.23-1.34)	0.26

**Supplementary table 2.** Cox analysis of TUDD for Social functioning dimension with MCID  $\geq 5$  and MCID  $\geq 10$

	Number of patients	Number of events	Univariate analysis (n=425)		Multivariate analysis (n=426)	
			HR (95% CI)	p-value	HR (95% CI)	p-value
120						
<b>Sex</b>						
Male	315		1.00			
Female	110		1.12 (0.76-1.66)	0.56		
<b>Age (years)</b>						
< 77	216		1.00			
$\geq 77$	209		1.06 (0.74-1.53)	0.74		
<b>Treatment</b>						
Monotherapy	213		1.00		1.00	
Doublet chemotherapy	212		0.73 (0.51-1.05)	0.09	0.71 (0.49-1.02)	0.06
<b>Performance status score</b>						
0-1	310		1.00			
2	114		1.29 (0.84-1.99)	0.25		
<b>Smoking status</b>						
Never smoked	89		1.00			
Ever smoked	336		1.14 (0.74-1.75)	0.56		
<b>Disease stage</b>						
III	82		1.00			
IV	343		1.02 (0.65-1.60)	0.93		
<b>Histology</b>						
Adenocarcinoma	215		1.00			
Squamous	145		1.26 (0.83-1.90)			
Other	65		1.09 (0.66-1.80)	0.53		
<b>MMS</b>						
$\leq 20$	32		1.00			
$> 20$	389		1.13 (0.50-2.57)	0.77		
<b>ADL</b>						
< 6	82		1.00			
6	335		0.93 (0.57-1.51)	0.77		
<b>CCI</b>						
$\leq 2$	317		1.00		1.00	
$> 2$	108		1.77 (1.17-2.70)	0.0073	1.82 (1.20-2.78)	0.0052
<b>BMI</b>						
$\leq 20$	47		1.00			
$20 < \text{BMI} \leq 30$	328		0.81 (0.47-1.40)			
$> 30$	50		0.54 (0.25-1.16)	0.27		

## **La qualité de vie relative à la santé comme co-critère de jugement principal dans les essais cliniques randomisés en oncologie**

### **Résumé**

La SG est le critère de référence mais avec l'augmentation du nombre de traitements efficaces disponibles dans beaucoup de cancers, il est nécessaire d'inclure un nombre plus important de patients et de les suivre plus longtemps ce qui augmente le coût des essais cliniques. Ainsi, les critères de jugement dit intermédiaires centrés sur la tumeur qui sont évalués plus précocement et utilisés comme critères substitutifs de la survie globale sont de plus en plus utilisés mais la plupart d'entre eux ne sont pas validés comme critères de substitution selon la méthodologie de référence. Ainsi la QdV constitue un critère de jugement pour évaluer un bénéfice clinique direct pour le patient. Un design alternatif serait d'associer un critère de jugement intermédiaire composite avec la QdV en tant que co-critères de jugement principaux. Dans cet article nous discutons les problématiques méthodologiques soulevées par un tel design : les règles de décision, les procédures de contrôle de l'erreur de type I, le calcul du nombre de sujets nécessaires.

EXPERT  
REVIEWSHealth-related quality-of-life  
as co-primary endpoint in  
randomized clinical trials in  
oncology*Expert Rev. Anticancer Ther.* Early online, 1–7 (2015)Frédéric Fiteni\*<sup>1,2</sup>,  
Alhousseiny Pam<sup>1</sup>,  
Amélie Anota<sup>1,3,4</sup>,  
Dewi Vernerey<sup>1</sup>,  
Sophie Paget-Bailly<sup>1</sup>,  
Virginie Westeel<sup>5</sup> and  
Franck Bonnetain<sup>1,3,4,6</sup><sup>1</sup>University Hospital of Besançon,  
Methodology and Quality of Life in  
Oncology Unit, Besançon, France<sup>2</sup>Department of Medical Oncology,  
University Hospital of Besançon,  
Besançon, France<sup>3</sup>EA 3181 University of Franche-Comté,  
Besançon, France<sup>4</sup>The French National Platform Quality  
of Life and Cancer, Besançon, France<sup>5</sup>Chest Disease Department, University  
Hospital of Besançon, Besançon, France<sup>6</sup>EORTC QOL Group, Brussels, Belgium

\*Author for correspondence:

Tel.: + 33 3 81 21 88 97

fredericfiteni@gmail.com

Overall survival (OS) has been considered as the most relevant primary endpoint but trials using OS often require large numbers of patients and long-term follow-up. Therefore composite endpoints, which are assessed earlier, are frequently used as primary endpoint but suffer from important limitations specially a lack of validation as surrogate of OS. Therefore, Health-related quality of life (HRQoL) could be considered as an outcome to judge efficacy of a treatment. An alternative approach would be to combine HRQoL with composite endpoints as co-primary endpoint to ensure a clinical benefit for patients of a new therapy. The decision rules of such design, the procedure to control the Type I error and the determination of sample size remain questions to debate. Here, we discuss HRQoL as co-primary endpoints in randomized clinical trials in oncology and provide some solutions to promote such design.

**KEYWORDS:** clinical trial • composite endpoint • co-primary endpoint • endpoint • health-related quality of life • methodology

Endpoints refer to clinical and biological measurements that assess the efficacy of therapeutic strategies. As the American Society of Clinical Oncology stated, active treatment in cancer is generally undertaken with the goal of providing improved quantity and/or quality of patient survival [1]. Cancer randomized clinical trials are conducted to obtain clinical evidence on the safety and efficacy of new interventions. The selection of an appropriately valid primary endpoint is an important aspect of clinical trial design to achieve this objective. Endpoints in cancer clinical trials can be classified into two main categories: patient-centered clinical endpoints, including overall survival (OS) and health-related quality of life (HRQoL), and tumor-centered clinical endpoints, such as progression-free survival (PFS) [2,3]. The US FDA considers OS benefit as the foundation for the approval of new anticancer drugs in the USA. Nevertheless, the increasing number of effective salvage treatments available in many types of cancer (i.e. subsequent lines of treatments) has resulted in the need for a larger number of patients to be included and/or the need of a more prolonged observation

period to attain sufficient events that can achieve planned statistical power; this increases the cost of clinical trials and requires a longer duration to obtain results [45]. Consequently, tumour-centered clinical endpoints such as PFS are often used as primary endpoints because they are assessed earlier (i.e., intermediate endpoints). These endpoints are frequently composite endpoints. However, there is a lack of consistency in their definitions, and they are not systematically validated as surrogate endpoints for OS [4]. In this context, HRQoL could constitute an alternative endpoint, which ensures earlier assessment of direct clinical benefit for the patient [6] and is recognized as a component endpoint for cancer therapy approvals by the American Society of Clinical Oncology and the FDA. An extract from the Lancet Handbook of essential concepts in clinical research says that “Researchers should restrict the number of primary endpoints tested. They should specify *a priori* the primary endpoint or endpoints in their protocol” but a new methodological approach would be to combined HRQoL with other endpoints, for example PFS as co-primary

endpoints, especially when OS is difficult to assess, to ensure the analysis of the clinical benefit of a new therapeutic. This paper discusses HRQoL as co-primary endpoints in randomized clinical trials in oncology.

### Composite endpoints

'Tumor-centered endpoints' are, most of the time, composite endpoints based on tumor assessment and combine different events such as local and distant progression, local and distant recurrence, development of metachronous cancer, death or severe toxicity. Composite endpoints combine multiple events (called components) into a single endpoint. An event is said to occur if any one of the prospectively defined components of the composite occurs. In the absence of at least one component considered in the composite endpoint, the patient is censored at the time of the last follow-up. For example, in the metastatic setting, PFS combined two components: tumor progression and death for any causes.

These types of endpoints are frequently used as primary outcome in oncologic clinical trials for several reasons. First, they can increase statistical power. A higher event rate than with OS is observed, so clinical trials need fewer numbers of patients to achieve required power [7]. Then they are assessed earlier than OS. The fewer number of subjects and the smaller duration of trials contribute to an economic benefit of less costly trials [89]. Then the use of a composite endpoint could lead to information preservation and thus reduction of bias due to information censoring [7]. For instance, PFS include death with a nonfatal event (tumor progression). In the case of a cancer with a poor prognosis, drop-out of the patient before the planned end of the study could correspond to patient's health deterioration. The censoring in this context seems to be informative; therefore, the inclusion of a nonfatal event reduces this bias. Moreover, the use of PFS may be appropriate in particular disease settings and has been accepted by regulatory authorities on many occasions [10,11].

Nevertheless, composite endpoints suffer from important limitations:

- There is heterogeneity of the definitions of composite endpoints like PFS, disease-free survival (DFS), time-to-treatment failure and so on. Consequently the definition of the same 'endpoint' is variable in different studies of the same disease [12]. Birgisson *et al.* [13] demonstrated that the inclusion of a second primary cancer other than the incident colorectal cancer as an event in the definition of DFS significantly impacted the results. The estimated DFS rate for patients with stage I–III disease was 62% after 5 years if this event was not counted as an event, compared with 58% if it was. The difference was larger for stage II (68 versus 60%) than for stage III (49 vs 47%). Another example is the PETACC 03 randomized study [14] where results were either significant or nonsignificant depending on whether or not second primary tumors were accounted for in the DFS definition. In the context of heterogeneity of definitions of survival endpoints, the international DATECAN initiative

(Definition for the Assessment of Time-to-event Endpoints in CANcer trials) [15] was developed to obtain standardized consensus definitions for multiple cancer sites based on a formal consensus process and has already led to guidelines for time-to-event endpoint definitions in sarcomas and gastrointestinal stromal tumors and pancreatic cancers trials [16,17].

- Most of composite endpoints have not been validated as a surrogate endpoint for OS. Nevertheless, PFS is frequently used as the primary outcome in other types of cancer, and many trials have demonstrated improved PFS with no improvement of OS. Currently, in cancer clinical trials, PFS was validated as a surrogate for OS only in 5-fluorouracil-based chemotherapy in advanced colorectal cancer [18], in T3/T4 rectal cancer [19].
- If an important component is not substantially modified by the effects of treatment, then less rather than more statistical power to detect effects may be the result [7].
- Although the composite endpoint as a whole may appear to be affected by the treatment, the benefit may be different between the components. This is the case in rectal cancer, and neoadjuvant chemo-radiotherapy has a higher effect on local recurrence than in metastatic recurrence [19].
- Competing risks interfere with the ability to observe events of primary interest in relation to the therapy being tested. Examples of competing risks can be second malignancies, treatment-related adverse effects. Such events interfere with analysis of effects on disease-specific events, which are the events directly affected by cancer therapies.

### HRQoL as co-primary endpoint

HRQoL constitutes an alternative endpoint that allows earlier assessment of direct clinical benefit for the patient. Over the last decade, a tremendous growth in the incorporation of HRQoL in cancer clinical trials has been observed. The FDA considers HRQoL as an endpoint that assesses direct clinical benefit for the patient [20,21]. In assessing new therapies and approaches, recent authors have first published the results of efficacy and elsewhere the analysis of HRQoL; for example, in breast cancer: BOLERO-2 [22,23], EMILIA [24,25], CLEOPATRA [26,27] trials, in lung cancer: LUX-lung 3 [28,29], OPTIMAL [30,31], AVAPERL [32,33] trials, in ovarian cancer: AURELIA [34,35], GOG 0218 [36,37] trials, in pancreatic cancer: PRODIGE 4/ACCORD 11 [38,39] trial. This indicates that HRQoL becomes a major consideration in treatment choice and influences adoption of new treatments. HRQoL could be implemented as composite primary endpoint (in combination with efficacy endpoints). Nevertheless, this approach might lead to problems of composite endpoints. This could develop new endpoints with heterogeneity between trials. The benefit may be different between the components; therefore the clinical meaningful of the endpoint would be difficult to interpret. Competing risks could interfere with the ability to observe events of primary interest in relation to the therapy being tested. It would be difficult for clinicians to interpret the

treatment effect. For example, in the AVAglio trial [40] that investigated the impact of adding bevacizumab to the standard therapy of newly diagnosed glioblastomas, deterioration-free survival was a composite endpoints defined as the time to: a  $\geq 10$ -point deterioration (regarded as clinically meaningful) from baseline without a subsequent  $\geq 10$ -point improvement from baseline; or progression or death. An exploratory sensitivity analysis was conducted without progression as an event. In the prespecified primary analysis, deterioration-free survival was significantly longer among patients in the bevacizumab group than among those in the placebo group for all five prespecified dimensions and all 21 nonprespecified dimensions, while in the sensitivity analysis, deterioration-free survival was significantly longer for only three of the five prespecified dimensions and 9 of 21 nonprespecified dimensions.

Another approach could be to combine intermediate endpoint, with HRQoL as co-primary endpoints, especially when treatments are associated with significant toxicity and surrogacy of other traditional efficacy clinical endpoints has not been demonstrated. A trial with co-primary endpoints is a trial with two or more primary efficacy endpoints. These endpoints can be composite endpoints. For example, Hussain *et al.* [41] compared intermittent and continuous androgen deprivation in metastatic prostate cancer, the co-primary objectives were to assess whether intermittent therapy was noninferior to continuous therapy with respect to survival, and whether HRQoL differed between the groups. C Von Plessen *et al.* compared a policy of three versus six courses of platinum-based combination chemotherapy with regard to effects on HRQoL and survival in patients with stage IIIB or IV non-small cell lung cancers [42]. Nevertheless, this approach is recent and methodological research must be pursued to clearly define decision-making rules to promote such design.

First, two main statistical methods have been developed to longitudinally analyze HRQoL: the linear mixed model for repeated measure (LMM) and the time to HRQoL deterioration (TTD), and at this time, no guidelines for the longitudinal analysis have been proposed, which compromise comparison between trials [6]. For example, two recent Phase III trials investigating the impact of adding bevacizumab to the standard therapy of newly diagnosed glioblastomas have applied two different approaches (LMM and TTD) to analyze longitudinal HRQoL data [40,43]. The results are divergent and compromise the conclusion about clinical value of adding bevacizumab since OS was not improved. Therefore, the analysis of HRQoL requires improved standardization in the future. Hence, LMM and TTD could systematically be applied or data could be available for researches, in order to make alternative analyses allowing the comparisons between trials.

Then, the treatment effect on HRQoL might be confounded by subsequent treatment. Nevertheless, most of the time, HRQoL is measured until progression; therefore, in this case, HRQoL is not confounded by subsequent treatment.

Then, in oncology, many self-completion HRQoL questionnaires have been developed and validated, and The European

Organization for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire QLQ-C30 and the Functional Assessment of Cancer Therapy-General (FACT-G) are the most widely used cancer-specific instrument in Phase III clinical trials [44,45]. A challenge would be to promote, through cancer site and treatment modalities, guidelines for selecting the best questionnaires allowing for direct comparison of results across trials. As an example, two recent Phase III clinical trials, OPTIMAL and LUX-Lung 3, focused on patients with advanced non-small cell lung cancer with EGFR mutations and investigated an inhibitor of EGFR (erlotinib for OPTIMAL trial and afatinib for LUX-Lung 3 trial respectively) versus chemotherapy [28,30]. HRQoL was assessed by the FACT-L questionnaire in OPTIMAL trial while EORTC QLQ-C30 and LC13 module were used in LUX-Lung 3 trial. These two clinical trials could hardly be compared since EORTC and FACT questionnaires for lung cancer do not contain the same dimensions in regard to impact of lung cancer on HRQoL.

The decision rules of trials with HRQoL as co-primary endpoint remains a question to debate. If an intermediate endpoint and HRQoL are co-primary endpoints, we can presume that the investigator might want to give equal importance to the endpoints. Therefore, a decision rule could be that the intermediate composite endpoint must be positive (e.g., PFS) and at least one dimension of HRQoL must be positive without deterioration of the other. Another approach of decision rules could be that statistical significance is needed for the intermediate composite endpoint and all HRQoL dimensions. This procedure is called intersection-union test. The global null hypothesis can be expressed as the union of all single hypotheses regarding the individual endpoints; this union is tested versus the global alternative that can be expressed as the intersection of all single alternative hypotheses. Moreover, we know that there is general agreement concerning the multidimensional concept of HRQoL taking into account levels of physical, mental, social and patient satisfaction with treatment. Therefore, the improvement of only one dimension of HRQoL reaches the problem of the holistic sense of HRQoL. The inclusion of an overall HRQoL dimension that captures everything and the inclusion of more than one dimension of HRQoL might be systematically discussed. The choice of the co-primary endpoints and dimensions of HRQoL must be discussed between clinicians and methodologists, and the decision rules must be clearly described in the protocol.

In confirmatory clinical trials with multiple endpoints, the use of multiple test procedures is mandatory and CONSORT Statement [46,47] recommends a multiplicity adjustment in case of multiple testing. However, no procedure to control the Type I error was implemented in the trials that analyzed HRQoL as secondary endpoint. In biomedical clinical research, it is well known that testing multiple hypotheses without any adjustment may increase the probability of erroneously rejecting at least one true null hypothesis. In an intersection-union test procedure, as all individual null hypotheses have to be rejected to claim significance, no multiplicity adjustment, that is, no

special method to consider the multiplicity of tests, is needed. Nevertheless, the Type II error rate inflates, that is, there is an increased probability of not rejecting a false null hypothesis [48]. This inflation must be taken into account for sample size determination. If the decision rules are not an intersection-union test procedure, the Type I error rate increases substantially with the total number of outcomes analyzed. Many procedures have been developed to control the Type I error at 0.05. Two main approaches can be distinguished: single-step procedures and stepwise procedures [48]. In a single-step procedure, such each single test is tested without reference to any other. The simplest approach to controlling such risk is the Bonferroni correction. In this procedure, the significance level  $\alpha$  is split equally among the  $k$  hypotheses, and each is tested at level  $\alpha/k$ . The individual null hypothesis  $H(i)$  is rejected if  $p\text{-value} < \alpha/k$ , and the global null hypothesis  $H$  is rejected if  $\min p\text{-value} < \alpha/k$ . The Bonferroni procedure is conservative and consequently leads to a diminution of power if the number of hypotheses  $k$  is large. As an illustration of this approach, Hussain *et al.* [41] used such design, the co-primary objectives were to assess whether intermittent therapy was noninferior to continuous therapy with respect to survival, and whether HRQoL differed between the groups. The five targeted HRQoL dimensions (impotence or erectile dysfunction, libido, vitality, mental health and physical functioning) were clearly specified in the protocol and the  $p$ -value of 0.01 was chosen for HRQoL to control the overall Type I error rate at 0.05. In stepwise procedures, such as testing in a hierarchical order, the decision on already tested hypotheses influences whether subsequent tests are conducted. The procedure is carried out sequentially where  $H(i)$  is tested at level  $\alpha$  as long as all previous hypotheses  $H(1), \dots, H(i-1)$  are rejected; testing stops with the first non-significant result. Holm developed a sequential procedure that improved upon the Bonferroni method. The univariate  $p$ -values are classified in increasing order such that  $P(1) \leq P(2) \leq \dots \leq P(k)$  with corresponding hypotheses  $H(1), H(2), \dots, H(k)$ . Then, the smallest  $p$ -value,  $P(1)$ , is compared against the most conservative  $\alpha$ -level,  $\alpha/k$ . If  $P(1) \leq \alpha/k$  holds, reject  $H(1)$  and continue stepwise, with each successive test conducted at progressively higher significance levels,  $\alpha/(k-1), \alpha/(k-2), \dots, \alpha/l$ . Otherwise, fail to reject  $H(1)$ , and stop the procedure. The acceptance of  $H(i)$  implies acceptance of  $H(j)$ , for all  $j > i$ . It is possible to assign different weights to the endpoints according to their importance within the Holm procedure. This method provides a substantial increase in power than the Bonferroni method, but a major problem of the sequential procedure is that once a hypothesis is not rejected, no further testing is permitted. The Bonferroni technique is the most classical but is conservative. This conservatism and the accompanied lack of power when the endpoints are correlated is the major drawback of the Bonferroni procedure. Holm's weighted sequential Bonferroni procedure gives the option to weight the individual hypotheses and proportionally allocate the  $p$ -level to account for the relative 'importance' of the individual endpoint [48]. Nevertheless, if an intermediate endpoint and HRQoL are

co-primary endpoints, we can presume that the investigator might want to give equal importance to these two endpoints. The weighted procedure might be rather discussed for the different dimensions of HRQoL related to their clinical importance. Moreover, these procedures lead to a reduction of power especially when the number of tests increases. Therefore, the greater the number of co-primary endpoints, the more stringent the alpha level becomes for each endpoint after adjustment and the sample size required increases. The determination of sample size, taking into account the adjustment of the type I error, is an urgent statistical challenge. One approach could be to repeat the sample size estimates for each endpoint and then select the largest number as the sample size required to answer all the questions of interest.

#### Expert commentary

The variety of composite time-to-event tumor-centered endpoints and the variability of their definitions are recognized as a major methodological problem and efforts must be pursued to improve their reliability and reproducibility. Moreover, most of composite endpoints have not been validated as a surrogate endpoint for OS and when evaluating a new class of therapy, trials must be performed to validate a surrogate endpoint; therefore the clinical benefit of these trials using such endpoints remains uncertain.

Consequently, HRQoL constitutes an alternative endpoint that allows earlier assessment of direct clinical benefit for the patient. A new approach is to combine intermediate endpoint, with HRQoL as co-primary endpoints, especially when treatments are associated with significant toxicity, and when surrogacy of other traditional efficacy clinical endpoints has not been demonstrated. Implementation of HRQoL as co-primary can provide a useful measure of the impact of intervention in a way that it is relevant to clinicians and patients especially when OS is difficult to measure.

#### Five-year view

In oncologic clinical trials, the Definition for the Assessment of Time-to-event Endpoints in CANcer trials (DATECAN) project has been developed to provide recommendations and to standardize definitions of time-to-event composite endpoints for each specific diseases and at each specific stage by use of a formal consensus methodology [15]. This is the preliminary step before assessing their surrogate capabilities (DATECAN 2 project).

HRQoL as co-primary endpoint is a recent approach, and methodological research must be pursued to promote such design. First, the measurement, analysis and reporting of HRQoL requires improved standardization in the future. Then, software packages which calculate the sample size in Phase III cancer clinical trials with HRQoL as co-primary endpoints should be developed where the investigators could specify the treatment effect for each co-primary and the procedure to adjust the type I error to obtain the sample size. Then, algorithms to clearly define decision-making rules of trials with HRQoL as co-primary endpoint must be developed.



**Financial & competing interests disclosure**

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This

includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

**Key issues**

- Composite endpoints combine multiple events (called components) into a single endpoint.
- There is heterogeneity of the definitions of composite endpoints like PFS.
- Most of composite endpoints have not been validated as a surrogate endpoint for OS. Nevertheless, PFS is frequently used as the primary outcome in other types of cancer and many trials have demonstrated improved PFS with no improvement of OS.
- A trial with co-primary endpoints is a trial with two or more primary efficacy endpoints.
- Over the last decade, a tremendous growth in the incorporation of HRQoL in cancer clinical trials has been observed. The US FDA considers HRQoL as an endpoint that assesses direct clinical benefit for the patient.
- Implementation of HRQoL as co-primary can provide a useful measure of the impact of intervention in a way that it is relevant to clinicians and patients.
- The decision rules of trials with HRQoL as co-primary endpoint remains a question to debate. If an intermediate endpoint and HRQoL are co-primary endpoints, we can presume that the investigator might want to give equal importance to the endpoints. Therefore, a decision rule could be that the intermediate composite endpoint must be positive (e.g., PFS) and at least one dimension of HRQoL must be positive without deterioration of the other. The decision rules must be clearly described in the protocol.
- The procedure to control the type I error must be clearly reported and the determination of sample size, taking into account the adjustment of the type I error, is mandatory.

**References**

Papers of special note have been highlighted as:

• of interest

•• of considerable interest

1. Peppercorn JM, Smith TJ, Helft PR, et al. American society of clinical oncology statement: toward individualized care for patients with advanced cancer. *J Clin Oncol* 2011;29(6):755-60
2. Booth CM, Ohorodnyk P, Eisenhauer EA. Call for clarity in the reporting of benefit associated with anticancer therapies. *J Clin Oncol* 2009;27(33):e213-14
3. Ohorodnyk P, Eisenhauer EA, Booth CM. Clinical benefit in oncology trials: is this a patient-centred or tumour-centred endpoint? *Eur J Cancer* 2009;45(13):2249-52
- **Description of how clinical benefit has been used in oncology trials.**
4. Fiteni F, Westeel V, Pivot X, et al. Endpoints in cancer clinical trials. *J Visc Surg* 2014;151(1):17-22
5. Saad ED, Buyse M. Overall survival: patient outcome, therapeutic objective, clinical trial end point, or public health measure? *J Clin Oncol* 2012;30(15):1750-4
6. 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
7. Sankoh AJ, Li H, D'Agostino RB. Use of composite endpoints in clinical trials. *Stat Med* 2014;33(27):4709-14
8. Freemantle N, Calvert M. Composite and surrogate outcomes in randomised controlled trials. *BMJ* 2007;334(7597):756-7
9. Freemantle N, Calvert M, Wood J, et al. Composite outcomes in randomized trials: greater precision but with greater uncertainty? *JAMA* 2003;289(19):2554-9
10. Blumenthal GM, Scher NS, Cortazar P, et al. First FDA approval of dual anti-HER2 regimen: pertuzumab in combination with trastuzumab and docetaxel for HER2-positive metastatic breast cancer. *Clin Cancer Res* 2013;19(18):4911-16
11. Thornton K, Kim G, Maher VE, et al. Vandetanib for the treatment of symptomatic or progressive medullary thyroid cancer in patients with unresectable locally advanced or metastatic disease: U.S. Food and Drug Administration drug approval summary. *Clin Cancer Res* 2012;18(14):3722-30
12. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, et al. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J Clin Oncol* 2008;26(22):3721-6
- **Evaluation of the reporting of survival end points in cancer randomized clinical trials. A majority of articles failed to provide a complete reporting of survival end points, thus adding another source of uncontrolled variability.**
13. Birgisson H, Wallin U, Holmberg L, et al. Survival endpoints in colorectal cancer and the effect of second primary other cancer on disease free survival. *BMC Cancer* 2011;11:438
14. Van Cutsem E, Labianca R, Bodoky G, et al. Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J Clin Oncol* 2009;27(19):3117-25
15. Bellera CA, Pulido M, Gourgou S, et al. 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* 2013;49(4):769-81
16. Bellera CA, Penel N, Ouali M, et al. Guidelines for time-to-event endpoint definitions in sarcomas and gastro-intestinal stromal tumors (GIST) trials. Results of the

- DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANCER trials). *Ann Oncol* 2015;26(5): 865-72
17. Bonnetain F, Bonsing B, Conroy T, et al. 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* 2014;50(17):2983-93
  18. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007;25(33):5218-24
  19. Bonnetain F, Bosset JF, Gerard JP, et al. What is the clinical benefit of preoperative chemoradiotherapy with 5FU/leucovorin for T3-4 rectal cancer in a pooled analysis of EORTC 22921 and FFCO 9203 trials: surrogacy in question? *Eur J Cancer* 2012; 48(12):1781-90
  20. Beitz J, Gnecco C, Justice R. Quality-of-life end points in cancer clinical trials: the U.S. Food and Drug Administration perspective. *J Natl Cancer* 1996(20):7-9
  - **The Oncologic Drugs Advisory Committee (US FDA) has recommended that beneficial effects on quality of life and/or survival be the basis for approval of new anticancer drugs. The FDA welcomes the opportunity to explore with investigators the use of quality of life instruments in the design of cancer clinical trials.**
  21. Johnson JR, Temple R. Food and drug administration requirements for approval of new anticancer drugs. *Cancer Treat Rep* 1985;69(10):1155-9
  22. Piccart M, Hortobagyi GN, Campone M, et al. Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2†. *Ann Oncol* 2014;25(12):2357-62
  23. Burris HA 3rd, Lebrun F, Rugo HS, et al. Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. *Cancer* 2013; 119(10):1908-15
  24. Verma S, Miles D, Gianni L, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med* 2012;367(19):1783-91
  25. Welslau M, Diéras V, Sohn JH, et al. Patient-reported outcomes from EMILIA, a randomized phase 3 study of trastuzumab emtansine (T-DM1) versus capecitabine and lapatinib in human epidermal growth factor receptor 2-positive locally advanced or metastatic breast cancer. *Cancer* 2014; 120(5):642-51
  26. Baselga J, Cortés J, Kim SB, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med* 2012;366(2):109-19
  27. Cortés J, Baselga J, Im YH, et al. Health-related quality-of-life assessment in CLEOPATRA, a phase III study combining pertuzumab with trastuzumab and docetaxel in metastatic breast cancer. *Ann Oncol* 2013;24(10):2630-5
  28. Wu YL, Zhou C, Hu CP, et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol* 2014;15(2):213-22
  29. Yang JC, Hirsh V, Schuler M, et al. Symptom control and quality of life in LUX-Lung 3: a phase III study of afatinib or cisplatin/pemetrexed in patients with advanced lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013;31(27): 3342-50
  30. Zhou C, Wu YL, Chen G, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* 2011;12(8):735-42
  31. Chen G, Feng J, Zhou C, et al. Quality of life (QoL) analyses from OPTIMAL (CTONG-0802), a phase III, randomised, open-label study of first-line erlotinib versus chemotherapy in patients with advanced EGFR mutation-positive non-small-cell lung cancer (NSCLC). *Ann Oncol* 2013;24(6): 1615-22
  32. Barlesi F, Scherpereel A, Rittmeyer A, et al. Randomized phase III trial of maintenance bevacizumab with or without pemetrexed after first-line induction with bevacizumab, cisplatin, and pemetrexed in advanced nonsquamous non-small-cell lung cancer: AVAPERL (MO22089). *J Clin Oncol* 2013;31(24):3004-11
  33. Rittmeyer A, Gorbunova V, Vikström A, et al. Health-related quality of life in patients with advanced nonsquamous non-small-cell lung cancer receiving bevacizumab or bevacizumab-plus-pemetrexed maintenance therapy in AVAPERL (MO22089). *J Thorac Oncol* 2013;8(11):1409-16
  34. Pujade-Lauraine E, Hilpert F, Weber B, et al. Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *J Clin Oncol* 2014;32(13):1302-8
  35. Stockler MR, Hilpert F, Friedlander M, et al. Patient-reported outcome results from the open-label phase III AURELIA trial evaluating bevacizumab-containing therapy for platinum-resistant ovarian cancer. *J Clin Oncol* 2014;32(13):1309-16
  36. Burger RA, Brady MF, Bookman MA, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. *N Engl J Med* 2011;365(26):2473-83
  37. Djärv T, Metcalfe C, Avery KN, et al. Prognostic value of changes in health-related quality of life scores during curative treatment for esophagogastric cancer. *J Clin Oncol* 2010;28(10):1666-70
  38. Conroy T, Desseigne F, Ychou M, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N Engl J Med* 2011;364(19):1817-25
  39. 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* 2013;31(1):23-9
  40. Chinot OL, Wick W, Mason W, et al. Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N Engl J Med* 2014;370(8):709-22
  41. Hussain M, Tangen CM, Berry DL, et al. Intermittent versus continuous androgen deprivation in prostate cancer. *N Engl J Med* 2013;368(14):1314-25
  42. Von Plessen C, Bergman B, Andresen O, et al. Palliative chemotherapy beyond three courses conveys no survival or consistent quality-of-life benefits in advanced non-small-cell lung cancer. *Br J Cancer* 2006;95(8):966-73
  43. Gilbert MR, Dignam JJ, Armstrong TS, et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N Engl J Med* 2014;370(8):699-708
  44. 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):365-76

- DATECAN initiative (Definition for the Assessment of Time-to-event Endpoints in CANcer trials). *Ann Oncol* 2015;26(5): 865-72
17. Bonnetain F, Bonsing B, Conroy T, et al. 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* 2014;50(17):2983-93
  18. Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007;25(33):5218-24
  19. Bonnetain F, Bosset JF, Gerard JP, et al. What is the clinical benefit of preoperative chemoradiotherapy with 5FU/leucovorin for T3-4 rectal cancer in a pooled analysis of EORTC 22921 and FFCD 9203 trials: surrogacy in question? *Eur J Cancer* 2012; 48(12):1781-90
  20. Beitz J, Gnecco C, Justice R. Quality-of-life end points in cancer clinical trials: the U.S. Food and Drug Administration perspective. *J Natl Cancer* 1996;20(7):9
  - **The Oncologic Drugs Advisory Committee (US FDA) has recommended that beneficial effects on quality of life and/or survival be the basis for approval of new anticancer drugs. The FDA welcomes the opportunity to explore with investigators the use of quality of life instruments in the design of cancer clinical trials.**
  21. Johnson JR, Temple R. Food and drug administration requirements for approval of new anticancer drugs. *Cancer Treat Rep* 1985;69(10):1155-9
  22. Piccart M, Hortobagyi GN, Campone M, et al. Everolimus plus exemestane for hormone-receptor-positive, human epidermal growth factor receptor-2-negative advanced breast cancer: overall survival results from BOLERO-2†. *Ann Oncol* 2014;25(12):2357-62
  23. Burris HA 3rd, Lebrun F, Rugo HS, et al. Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. *Cancer* 2013; 119(10):1908-15
  24. Verma S, Miles D, Gianni L, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med* 2012;367(19):1783-91
  25. Welslau M, Diéras V, Sohn JH, et al. Patient-reported outcomes from EMILIA, a randomized phase 3 study of trastuzumab emtansine (T-DM1) versus capecitabine and lapatinib in human epidermal growth factor receptor 2-positive locally advanced or metastatic breast cancer. *Cancer* 2014; 120(5):642-51
  26. Baselga J, Cortés J, Kim SB, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med* 2012;366(2):109-19
  27. Cortés J, Baselga J, Im YH, et al. Health-related quality-of-life assessment in CLEOPATRA, a phase III study combining pertuzumab with trastuzumab and docetaxel in metastatic breast cancer. *Ann Oncol* 2013;24(10):2630-5
  28. Wu YL, Zhou C, Hu CP, et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol* 2014;15(2):213-22
  29. Yang JC, Hirsh V, Schuler M, et al. Symptom control and quality of life in LUX-Lung 3: a phase III study of afatinib or cisplatin/pemetrexed in patients with advanced lung adenocarcinoma with EGFR mutations. *J Clin Oncol* 2013;31(27): 3342-50
  30. Zhou C, Wu YL, Chen G, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol* 2011;12(8):735-42
  31. Chen G, Feng J, Zhou C, et al. Quality of life (QoL) analyses from OPTIMAL (CTONG-0802), a phase III, randomised, open-label study of first-line erlotinib versus chemotherapy in patients with advanced EGFR mutation-positive non-small-cell lung cancer (NSCLC). *Ann Oncol* 2013;24(6): 1615-22
  32. Barlesi F, Scherpereel A, Rittmeyer A, et al. Randomized phase III trial of maintenance bevacizumab with or without pemetrexed after first-line induction with bevacizumab, cisplatin, and pemetrexed in advanced nonsquamous non-small-cell lung cancer: AVAPERL (MO22089). *J Clin Oncol* 2013;31(24):3004-11
  33. Rittmeyer A, Gorbunova V, Vikström A, et al. Health-related quality of life in patients with advanced nonsquamous non-small-cell lung cancer receiving bevacizumab or bevacizumab-plus-pemetrexed maintenance therapy in AVAPERL (MO22089). *J Thorac Oncol* 2013;8(11):1409-16
  34. Pujade-Lauraine E, Hilpert F, Weber B, et al. Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *J Clin Oncol* 2014;32(13):1302-8
  35. Stockler MR, Hilpert F, Friedlander M, et al. Patient-reported outcome results from the open-label phase III AURELIA trial evaluating bevacizumab-containing therapy for platinum-resistant ovarian cancer. *J Clin Oncol* 2014;32(13):1309-16
  36. Burger RA, Brady MF, Bookman MA, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. *N Engl J Med* 2011;365(26):2473-83
  37. Djärv T, Metcalfe C, Avery KN, et al. Prognostic value of changes in health-related quality of life scores during curative treatment for esophagogastric cancer. *J Clin Oncol* 2010;28(10):1666-70
  38. Conroy T, Desseigne F, Ychou M, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N Engl J Med* 2011;364(19):1817-25
  39. 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* 2013;31(1):23-9
  40. Chinot OL, Wick W, Mason W, et al. Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N Engl J Med* 2014;370(8):709-22
  41. Hussain M, Tangen CM, Berry DL, et al. Intermittent versus continuous androgen deprivation in prostate cancer. *N Engl J Med* 2013;368(14):1314-25
  42. Von Plessen C, Bergman B, Andresen O, et al. Palliative chemotherapy beyond three courses conveys no survival or consistent quality-of-life benefits in advanced non-small-cell lung cancer. *Br J Cancer* 2006;95(8):966-73
  43. Gilbert MR, Dignam JJ, Armstrong TS, et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N Engl J Med* 2014;370(8):699-708
  44. 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):365-76

## **Valeur pronostique de la qualité de vie relative à la santé sur la survie globale chez les patients âgés atteints de CBNPC**

### **Résumé**

#### **Contexte**

Nous avons étudié si la QdV est pronostique de la SG chez des patients âgés atteints de CBNPC.

#### **Méthodes**

Au total, 451 patients âgés de 70-89 ans atteints de CBNPC avancés ont été inclus dans l'essai IFCT-0501. Le questionnaire QLQ-C30 administré à baseline a été utilisé pour étudier la valeur pronostique de la QdV sur la SG. Un modèle de Cox a été utilisé pour l'analyse univariée et multivariée.

#### **Résultats**

La dimension de santé globale (GH) à baseline est associées à une SG favorable après ajustement sur les facteurs cliniques, fonctionnels et histologique (hazard ratio [HR]: 0.986; IC 95% : 0.980-0.992).

Nous avons distingué 3 groupes selon le score GH : haut (GH <46), intermédiaire ( $46 \leq GH \leq 67$ ), et bas (GH >67) risque de mortalité. La médiane de SG était de 14,5, 8,2 et 5,3 mois dans les bas, intermédiaire et haut risque de mortalité respectivement (log-rank P <0.0001). Dans le groupe à haut risque, le doublet de chimiothérapie n'était pas associé à une survie favorable (HR: 0.70; IC 95% : 0.49-1.003; P=0.052), alors que dans le groupe intermédiaire et le groupe de bas risque, le doublet de chimiothérapie était associé avec une survie globale favorable (HR: 0.72; IC 95%: 0.54-0.96; P=0.023 et HR: 0.50; IC 95%: 0.30-0.84; P=0.0089, respectivement).

#### **Conclusion**

Cette étude démontre la valeur pronostique de la QdV au diagnostic pour identifier des sous-groupes de populations vulnérables chez les patients âgés atteints de CBNPC avancés. De plus, la QdV pourrait sélectionner les patients qui vont bénéficier du doublet de chimiothérapie.



Original Research

# Prognostic value of health-related quality of life for overall survival in elderly non-small-cell lung cancer patients



Frédéric Fiteni <sup>a,b,c,\*</sup>, Dewi Vernerey <sup>a,1</sup>, Franck Bonnetain <sup>a,c,d</sup>, Fabien Vaylet <sup>e</sup>, Hélène Sennéart <sup>f</sup>, Jean Trédaniel <sup>g</sup>, Denis Moro-Sibilot <sup>h</sup>, Dominique Herman <sup>i</sup>, Hélène Laizé <sup>j</sup>, Philippe Masson <sup>k</sup>, Marc Derollez <sup>l</sup>, Christelle Clément-Duchêne <sup>m</sup>, Bernard Milleron <sup>n</sup>, Franck Morin <sup>n</sup>, Gérard Zalzman <sup>n,o</sup>, Elisabeth Quoix <sup>n,p</sup>, Virginie Westeel <sup>n,q</sup>

<sup>a</sup> University Hospital of Besançon, Methodology and Quality of Life in Oncology Unit, Besançon, France

<sup>b</sup> University Hospital of Besançon, Department of Medical Oncology, Besançon, France

<sup>c</sup> EA 3181 University of Franche-Comté, Besançon, France

<sup>d</sup> Platform Quality of Life and Cancer, France

<sup>e</sup> Instruction Hospital of the Armies, Department of Pneumology, Percy-Clamart, France

<sup>f</sup> CLCC René Gauducheau, Department of Medical Oncology, Nantes, France

<sup>g</sup> St Joseph Hospital, Department of Medical Oncology, Paris, France

<sup>h</sup> University Hospital of Grenoble, Pneumology Department, Grenoble, France

<sup>i</sup> Hospital of Nevers, Pneumology Department, Nevers, France

<sup>j</sup> Hospital of Sceaux, Pneumology Department, Sceaux, France

<sup>k</sup> Hospital of Cholet, Pneumology Department, Cholet, France

<sup>l</sup> Private Hospital, Pneumology Department, Maubeuge, France

<sup>m</sup> University Hospital of Nancy, Pneumology Department, Nancy, France

<sup>n</sup> Intergroupe Francophone de Cancérologie Thoracique (IFCT), Paris, France

<sup>o</sup> University Hospital of Caen, Pneumology Department, Caen, France

<sup>p</sup> University Hospital of Strasbourg, Pneumology Department, France

<sup>q</sup> University Hospital of Besançon, Pneumology Department, France

Received 13 August 2015; received in revised form 20 September 2015; accepted 4 October 2015

Available online xxx

\* Corresponding author: University Hospital of Besançon, Department of Medical Oncology, 3 Boulevard Fleming, 25000, Besançon, France. Tel.: +333 81668796.

E-mail address: [fredericfiteni@gmail.com](mailto:fredericfiteni@gmail.com) (F. Fiteni).

<sup>1</sup> Shared authorship.

**KEYWORDS**

Quality of life;  
Prognostic factor;  
Methodology;  
Lung cancer

**Abstract Background:** We investigated whether the health-related quality of life (HRQoL) score is a prognostic factor for overall survival (OS) in elderly patients with advanced non-small-cell lung cancer (NSCLC).

**Methods:** We included 451 NSCLC patients aged 70–89 years enrolled in the Intergroupe Francophone de Cancérologie Thoracique 0501 trial, using scores of the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire Core 30 at baseline to investigate the prognostic value of HRQoL for OS, in addition to conventional factors. Cox regression model was used for both univariate and multivariate analyses of OS.

**Results:** Global health status (GH) dimension score at baseline was associated with favourable OS when adjusted for clinical, functional, and histological factors (hazard ratio [HR]: 0.986; 95% confidence interval [CI]: 0.980–0.992).

We distinguished three groups according to GH score: high (GH <46), intermediate (46 ≤ GH ≤ 67), and low (GH >67) mortality risk. The median OS values were 14.5, 8.2, and 5.3 months in the low-, intermediate-, and high-risk categories, respectively (log-rank P < 0.0001).

In the high-risk group, doublet chemotherapy was not associated with favourable OS (HR: 0.70; 95% CI: 0.49–1.003; P = 0.052), whereas in the intermediate- and low-risk groups, doublet chemotherapy was associated with favourable OS (HR: 0.72; 95% CI: 0.54–0.96; P = 0.023 and HR: 0.50; 95% CI: 0.30–0.84; P = 0.0089, respectively).

**Conclusion:** This study supports the additional prognostic value of HRQoL data at diagnosis to identify vulnerable subpopulations in elderly NSCLC patients. HRQoL could thus be valuable in selecting patients who will benefit from doublet chemotherapy.

© 2015 Elsevier Ltd. All rights reserved.

## 1. Background

The number of studies using health-related quality of life (HRQoL) assessment has been growing over the last decade. The Food and Drug Administration considers HRQoL to be an end-point for assessing direct clinical benefits for the patient [1–3]. Moreover, there has been evidence to suggest that assessing baseline HRQoL dimension scores in cancer patients improves the prediction of overall survival (OS) [4–9]. Quinten et al. carried out a meta-analysis involving over 10,000 cancer patients (16% lung cancer), revealing that baseline HRQoL was a prognosticator of longer survival [10]. In non-small-cell lung cancer (NSCLC) patients, several studies have demonstrated that HRQoL represents a significant prognosticator of favourable OS [6,9,11–13]. Sloan et al. prospectively observed 2,442 patients with stage I–IV NSCLC, 47% ≤ 65 years old and 53% > 65 years old, all completing a single-item measure of overall HRQoL from the Lung Cancer Symptom Scale questionnaire within the first 6 months post-diagnosis. They demonstrated that QoL deficits at diagnosis were significantly associated with shorter OS (hazard ratio [HR]: 1.55; P < 0.001). Yet no study has specifically focused on elderly advanced NSCLC patients.

We sought to investigate the additional prognostic value of baseline HRQoL assessed by European Organisation for Research and Treatment of Cancer (EORTC) Quality of Life Questionnaire Core 30 (QLQ-

C30) in elderly advanced NSCLC patients treated with chemotherapy in the randomised Intergroupe Francophone de Cancérologie Thoracique (IFCT) 0501 trial.

## 2. Methods

### 2.1. Sample

The IFCT 0501 study design has been previously described [14]. Patients aged 70–89 years with stage IV NSCLC or stage III unsuitable for radical radiation therapy and performance status (PS) ≤ 2 were eligible for this phase III trial. They were randomly assigned 1:1 to four 28-day cycles of monthly carboplatin plus weekly paclitaxel or five 21-day cycles of single agent vinorelbine or gemcitabine, on days 1 and 8 of each cycle. Patients were stratified by centre, World Health Organization (WHO) PS score (0–1 versus 2), stage (III versus IV), and age (≤ 80 versus > 80 years).

The protocol was approved by the *Comité de Protection des Personnes* of Ile-de-France X, Aulnay-sous-Bois, France, the trial being authorised by the French National Authority for Health. All patients provided written informed consent.

### 2.2. Health-related quality of life

HRQoL was assessed using EORTC QLQ-C30 questionnaire [15] at randomisation, then at 6 and 18 weeks.

The QLQ-C30 is a cancer-specific tool composed of 30 items [16–18]. Five functional scores (physical, role, cognitive, social, and emotional) have been developed, rated on a global health score ranging from 0 (worst) to 100 (best), as well as nine symptom scores (nausea, pain [PA], fatigue, dyspnoea, difficulty sleeping, anorexia, constipation, diarrhoea, and perceived financial difficulties), ranging from 0 (best) to 100 (worse).

### 2.3. Statistical analysis

We used means and standard deviations for continuous variables, and proportions for qualitative variables. We compared means and proportions using Student's t-test and chi-squared test, or Fisher's exact test if appropriate.

Patient characteristics were described based on the completion of baseline QoL questionnaire, ensuring that any non-random missing patient profiles were detected.

OS was defined as the time from randomisation to death from any cause. Survival was censored at last follow-up or time of analysis. OS was estimated using the Kaplan-Meier method and presented as median with 95% confidence interval (CI). Follow-up was calculated using a reverse Kaplan-Meier estimation.

The association of non-HRQoL characteristics at diagnosis and baseline HRQoL dimensions in terms of OS was assessed using univariate Cox regression analysis, followed by multivariate analysis for those exhibiting  $P < 0.1$ . The factors identified with a  $P < 0.1$  in multivariate analysis were thereafter included in a final multivariate model with stepwise backward elimination ( $P < 0.05$ ).

The proportionality assumption for the Cox model was verified using the log graphic method. Collinearity of baseline HRQoL scores with other covariates was examined using a multiple linear regression model.

The Hosmer and Lemeshow goodness-of-fit statistics test, adapted for survival analysis, was used to evaluate the final model's calibration.

Internal validation using a bootstrap procedure was performed to assess the final model's robustness, analysing hazard regression uncertainty for parameters involved in the final model [19].

The prognostic value of HRQoL scores added to a reference risk model, including the non-HRQoL characteristics enrolled in the final multivariate model, was evaluated using C-statistics. Harrell's C-index estimates discriminate capability, i.e. the ability to distinguish between high-risk and low-risk patients, the C-index varying from 0.5 (no discrimination) to 1 (perfect discrimination). This analysis was repeated 1000 times using bootstrap samples to derive 95% CIs for the between-model difference in C-statistics.

We used continuous net reclassification improvement (cNRI) and integrated discrimination

improvement (IDI) to quantify the performance and net benefit of adding HRQoL scores to the reference model at 24 months post-randomisation [20,21]. The cNRI quantifies the direction of change, and the IDI the magnitude of change. When significantly greater than 0, IDI and cNRI suggest the existence of a net benefit through adding the marker of interest to the reference model.

To implement HRQoL scoring into clinical practice, we determined a cut-off value via an unsupervised method using the Q1 and Q3 interquartiles.

We performed sensitivity analyses. We first included HRQoL and clinical factors with a  $P < 0.1$  in the univariate analysis using a stepwise backward elimination procedure. We then conducted a stepwise multivariate model with the treatment variable as stratification factor for the final model construction. As some data from the activities of daily living (ADL), mini-mental state (MMS), and HRQoL questionnaires could have been redundant by the time of analysis, we eventually conducted a stepwise multivariable model excluding the ADL score, then the MMS score, and eventually both.

Tests were two-sided, with P-values  $< 0.05$  considered significant. The analyses were conducted using SAS 9.2 (SAS, Cary, NC) and R software (Version 2.10.1).

## 3. Results

### 3.1. Study population

Between April 2006 and December 2009, 451 patients were enrolled. The number of patients who completed the entire questionnaire at baseline was 361 (80.04%), the number of available questionnaires (i.e. with at least one QoL score that could be calculated) being 421 (93.3%). At baseline, the patients who completed the entire QoL questionnaire and those who did not were found to display similar clinical characteristics (Table 1). The baseline HRQoL scores for each dimension have been presented in Table 2. Median follow-up was 30.3 months (range: 8.6–45.2). There were 199 (88%) deaths under monotherapy versus 177 (78.6%) under doublet chemotherapy.

### 3.2. Association between baseline HRQoL scores and overall survival

Data pertaining to the association of clinical and HRQoL dimension scores in terms of OS is shown in Tables 3A and B. In the final model, the following characteristics were independent favourable prognosticators of OS: increasing global health status (GH) score (HR: 0.986; 95% CI: 0.980–0.992;  $P < 0.0001$ ), PS 0–1 (HR: 0.63; 95% CI: 0.49–0.81;  $P < 0.0001$ ), doublet chemotherapy (HR: 0.65; 95% CI: 0.52–0.80;  $P$

**Table 1**  
Patient characteristics according to Quality of Life Questionnaire Core 30 completion.

	Patients who completed the questionnaire at baseline (n=361)		Patients who did not complete the questionnaire at baseline (n=90)		Fisher's exact test P-value
	n	(%)	n	(%)	
<b>Age</b>					
<77	180	49.9	48	53	0.56
≥77	181	50.1	42	47	
<b>Gender</b>					
Male	270	74.8	63	70	0.35
Female	91	25.2	27	30	
<b>Performance status</b>					
0–1	267	74	63	70	0.15
2	93	25.7	27	30	
Unknown	1	0.3	0		
<b>Smoking status</b>					
Never smoked	73	20	21	23	0.51
Ever smoked	288	80	69	77	
<b>MMS</b>					
≤20	29	8	6	7	0.82
>20	330	91	76	84	
Unknown	2	1	8	9	
<b>ADL</b>					
<6	66	18	22	24	0.12
6	288	80	62	69	
Unknown	7	2	6	7	
<b>CCI</b>					
≤2	268	74	73	81	0.17
>2	93	26	17	19	
<b>BMI</b>					
≤20	43	12	9	10	0.77
20<BMI≤30	276	76	72	80	
>30	42	12	9	10	
<b>Stage</b>					
IIIA–IIIB	70	19	17	19	0.9
IV	291	81	73	81	
<b>Histology</b>					
Adenocarcinoma	184	51	45	50	0.63
Squamous	123	34	28	21	
Other	54	15	17	19	

MMS = mini-mental state score; ADL = activities of daily living score; CCI = Charlson comorbidity index; BMI = body mass index.

<0.0001), never smoked status (HR: 0.58; 95% CI: 0.43–0.78; P=0.0003), adenocarcinoma (HR: 0.68; 95% CI: 0.50–0.93; P=0.047), increasing ADL score (HR: 0.73; 95% CI: 0.61–0.88; P=0.0011), and increasing MMS score (HR: 0.97; 95% CI: 0.94–0.99, P=0.044) (Table 4). The model's calibration was acceptable (Hosmer–Lemeshow with deciles p=0.1). The internal validation HR uncertainties reflected its robustness, especially the association between GH score and OS (Table 4).

Entering the GH score into the reference model was found to significantly improve its discriminative ability, notably its capacity to discriminate between patients who died and those who did not, as C-index

**Table 2**  
Health-related quality of life scores at baseline by treatment arm.

QLQ-C30 scores	All patients (n=451)		
	N*	Mean (SD)	Median (minimum–maximum)
Global health status	420	56.8 (18.7)	58.3 (8.3–100)
Physical functioning	420	69.2 (22.5)	73.3 (0–100)
Role functioning	420	66 (34.2)	66.7 (0–100)
Emotional functioning	420	72.2 (22.6)	75 (0–100)
Cognitive functioning	421	82.4 (21.1)	82.3 (0–100)
Social functioning	411	78 (30.5)	100 (0–100)
Fatigue	420	43.5 (27.8)	33.3 (0–100)
Nausea/vomiting	421	5.5 (15.2)	0 (0–100)
Pain	420	27.9 (29.8)	16.7 (0–100)
Dyspnoea	419	44.2 (34.7)	33.3 (0–100)
Insomnia	420	28.9 (32.3)	33.3 (0–100)
Appetite loss	419	35.8 (37.2)	33.3 (0–100)
Constipation	420	25.6 (32)	0 (0–100)
Diarrhoea	417	7.5 (18.4)	0 (0–100)
Financial problems	416	4.4 (15.2)	0 (0–100)

SD = standard deviation; QLQ-C30 = Quality of Life Questionnaire Core 30.

\* Number of patients with HRQoL score at baseline that can be calculated high score for a functional scale represents a *high/healthy level of functioning*, high score for the global health status/HRQoL represents a *high HRQoL*, but high score for a symptom scale/item represents a *high level of symptomatology/problems*.

statistics significantly increased from 0.66 to 0.69 (bootstrap mean difference: 0.0253; 95% CI: 0.0248–0.0259). Similarly, including the GH score into the reference model adequately reclassified patients into lower (no events) or higher (events) mortality risk, as demonstrated by a continuous net reclassification index of 0.38 (95% CI: 0.13–0.64) at 24 months post-randomisation. When adding the GH score adequately reclassified 42/68 patients (61.8%) into the 'no event' group, whereas it reclassified 196/341 patients (57.4%) into the 'event' group (Fig. 1). The IDI was 0.03 (P=0.0006). Adding the GH score to the classical risk factors improved the stratification of patients at risk of death.

The factors affecting baseline GH scores were explored using multiple linear regression. Increasing baseline GH score was associated with a PS=0–1 (P<0.0001), increasing ADL score (P=0.0061), and increasing body mass index (BMI) (P=0.023). Nevertheless, this model exhibited R<sup>2</sup> statistics of 12%, indicating that the GH score was not completely accounted for by PS, ADL score, and BMI.

### 3.3. Sensitivity analyses

In the multivariate Cox model including all variables, GH dimension score remained associated with OS (HR: 0.98; 95% CI: 0.97–0.99; P=0.0002) (Supplementary Table 1). All variables significantly influencing OS in the final model were also found to be statistically



Table 3A  
Univariate and multivariate Cox regression with clinical parameters associated with overall survival.

	Number of patients	Number of events	Univariate analysis (n=451)		Multivariate analysis (n=393)	
			HR (95% CI)	P-value	HR (95% CI)	P-value
Gender						
Male	333	282	1.00		1.00	
Female	118	95	0.77 (0.61–0.97)	0.026	1.053 (0.76–1.47)	0.76
Age (years)	451	377	0.99 (0.97–1.02)	0.60		
Treatment						
Monochemotherapy	225	199	1.00		1.00	
Doublet chemotherapy	226	178	0.64 (0.52–0.78)	<0.0001	0.61 (0.50–0.77)	<0.0001
Performance status score						
0–1	327	262	1.00		0.59 (0.46–0.75)	
2	123	114	2.10 (1.67–2.60)	<0.0001	1.00	<0.0001
Smoking status						
Never smoked	87	68	0.65 (0.50–0.84)		0.58 (0.39–0.85)	
Ever smoked	364	309	1.00	0.001	1.00	0.0034
Disease stage						
III	82	71	1.00			
IV	349	306	1.05 (0.81–1.36)	0.71		
Histology						
Adenocarcinoma	219	178	0.55 (0.41–0.73)		0.65 (0.48–0.87)	
Squamous	146	131	0.75 (0.56–1.01)		0.76 (0.56–1.03)	
Other	66	68	1.00	<0.0001	1.00	0.0015
MMS	441	377	0.96 (0.93–0.98)	0.0013	0.97 (0.94–1.00)	0.057
ADL	438	377	0.64 (0.55–0.74)	<0.0001	0.70 (0.59–0.83)	<0.0001
CCI	451	377	1.06 (0.99–1.13)	0.086	0.98 (0.96–1.01)	0.77
BMI	451	377	0.98 (0.95–1.00)	0.067	1.01 (0.94–1.09)	0.17

HR = hazard ratio; CI = confidence interval; MMS = mini-mental state; ADL = activities of daily living; CCI = Charlson comorbidity index; BMI = body mass index.

significant in this model. Only one variable, not included into the final model, significantly and negatively correlated with OS: the PA dimension score (HR: 1.004; 95% CI: 1.00–1.01; P=0.046).

In the model including treatment as stratification variable, GH score, PS 0–1, never smoked status, ADL score, and MMS score were favourably associated with OS (Supplementary Table 2).

GH score remained statistically significant in all sensitivity analyses, conducted without ADL, MMS and both scores. All other covariates significantly associated with OS in our final model were found to be statistically significant prognosticators of OS (Supplementary Tables 3–5). Physical functioning (PF) score was favourably associated with OS when the ADL score was not included.

Table 3B  
Univariate and multivariate Cox regression only with health-related quality of life dimensions scores associated with overall survival.

	Number of patients	Number of events	Univariate analysis (n=451)		Multivariate analysis (n=451)	
			HR (95% CI)	P-value	HR (95% CI)	P-value
Global health status	420	377	0.98 (0.98–0.99)	<0.0001	0.986 (0.98–0.992)	<0.0001
Physical functioning	420	377	0.98 (0.98–0.99)	<0.0001	0.989 (0.984–0.995)	0.0003
Role functioning	420	377	0.99 (0.987–0.993)	<0.0001	0.994 (0.991–0.998)	0.001
Emotional functioning	420	377	0.989 (0.989–0.998)	0.0055	0.996 (0.991–1.000)	0.068
Cognitive functioning	421	377	0.992 (0.987–0.997)	0.0011	0.996 (0.991–1.001)	0.12
Social functioning	411	377	0.993 (0.989–0.996)	<0.0001	0.997 (0.993–1.001)	0.12
Fatigue	420	377	1.011 (1.007–1.015)	<0.0001	1.007 (1.003–1.011)	0.0012
Nausea and vomiting	421	377	1.003 (0.996–1.010)	0.40		
Pain	420	377	1.010 (1.006–1.013)	<0.0001	1.007 (1.003–1.011)	0.0003
Dyspnoea	419	377	1.007 (1.004–1.010)	<0.0001	1.004 (1.000–1.007)	0.031
Insomnia	420	377	1.003 (1.000–1.006)	0.0983	1.002 (0.998–1.005)	0.33
Appetite loss	419	377	1.007 (1.004–1.010)	<0.0001	1.005 (1.001–1.008)	0.006
Constipation	420	377	1.004 (1.000–1.007)	0.025	1.002 (0.999–1.006)	0.22
Diarrhoea	417	377	1.004 (0.998–1.009)	0.23		
Financial difficulties	416	377	1.005 (0.998–1.012)	0.14		

HR = hazard ratio; CI = confidence interval.

Table 4  
Clinical and health-related quality of life dimension scores associated with overall survival.

	Number of patients	Number of events	HR (95% CI)	P-value
GH	420	377	0.986 (0.980–0.992)	<0.0001
Treatment				
Monochemotherapy	225	199	1.00	
Doublet chemotherapy	226	198	0.65 (0.52–0.80)	<0.0001
Performance status score				
0–1	327	262	0.63 (0.49–0.81)	
2	123	114	1.00	0.0003
Smoking status				
Never smoked	87	68	0.58 (0.43–0.78)	
Ever smoked	264	309	1.00	0.0003
Histology				
Adenocarcinoma	219	178	0.68 (0.50–0.93)	
Squamous	146	131	0.80 (0.58–1.01)	
Other	66	68	1.00	0.047
MMS	441	377	0.97 (0.94–1.00)	0.043
ADL	438	377	0.81 (0.66–0.99)	0.0011

HR = hazard ratio; CI = confidence interval; GH = global health score; MMS = mini-mental state; ADL = activities of daily living.

### 3.4. Proposal for implementing HRQoL

When using the interquartile ranges ( $GH \leq 46$ ,  $46 < GH < 67$ , and  $GH \geq 67$ ), GH score remained associated with OS in stepwise multivariable Cox regression ( $P < 0.0001$ ) (Table 5). We thus distinguished three groups, categorised as high ( $GH < 46$ ), intermediate ( $46 \leq GH \leq 67$ ), or low ( $GH > 67$ ) mortality risk. Median OS values were 5.3, 8.2, and 14.5 months in the low-, intermediate-, and high-risk groups, respectively (log-rank  $P < 0.0001$ ) (Fig. 2).

In the high-risk subgroup, doublet chemotherapy was not associated with favourable OS (HR: 0.70; 95% CI: 0.49–1.003;  $P = 0.052$ ). In the intermediate- and low-risk groups, doublet chemotherapy was associated with favourable OS (HR: 0.72; 95% CI: 0.54–0.96;  $P = 0.023$ , and HR: 0.50; 95% CI: 0.30–0.84;  $P = 0.0089$ , respectively).

## 4. Discussion

To our knowledge, this is the first analysis of HRQoL data derived from the EORTC QLQ-C30 questionnaire as prognostic markers of OS in elderly advanced NSCLC patients. Based on our data, the GH dimension score provided significant value in addition to PS, treatment type, smoking status, histology, and both ADL and MMS scores.

This is in line with other studies investigating HRQoL in NSCLC patients [6,14–16]. Sloan et al. [12] and Jacot et al. [13] demonstrated that overall HRQoL deficits, at lung cancer diagnosis, were significantly associated with poor OS (HR: 1.55;  $P < 0.001$  and 2.20;  $P < 0.001$ , respectively). Yet these studies did not take account of disease stage or age.

Among the Charlson comorbidity index (CCI), MMS score, and ADL score, no geriatric index has been

found to be able to guide thoracic oncologists in decision-making for elderly NSCLC patients. In our final model including GH, the MMS and ADL scores were both associated with OS, whereas CCI was not, as previously-published [14]. Our results indicate that HRQoL by the EORTC QLQ-C30 questionnaire could provide a useful tool, the GH score being statistically significant in all sensitivity analyses. Moreover, with PF being statistically associated with OS in the model without ADL, these results suggest that the EORTC QLQ-C30 questionnaire could even surpass the ADL score, reflecting the same characteristics while adding the global GH evaluation. The collinearity of baseline HRQoL scores with other covariates was examined using a multiple linear regression model. This model exhibited  $R^2$  statistics of 12%, indicating that the GH score was much more than a simple amalgamation of PS, ADL score and BMI. Moreover, the PS and ADL scores are evaluated by the physician rather than being self-reported. Therefore, the use of the HRQoL questionnaire could limit the interpretation by the physicians. Finally, the idea would be to use only the HRQoL questionnaire rather than two or more questionnaires (ADL, MMS etc) for the geriatric evaluation to help physicians in their decision-making, which is often difficult in elderly patients with lots of comorbidities.

Subgroup analyses suggested the baseline GH score to be a predictor of treatment effect, with 46 being the cut-off value. The ESOGIA-Groupe Français de Pneumo-Cancérologie 0802 trial assessed the integration of the comprehensive geriatric assessment (CGA) in treatment-decision-making in stage IV NSCLC patients over 70 years old [22]. The study failed to prove the superiority of a CGA-based strategy compared to PS-guided strategy of treatment allocation in terms of time to treatment failure. CGA has never

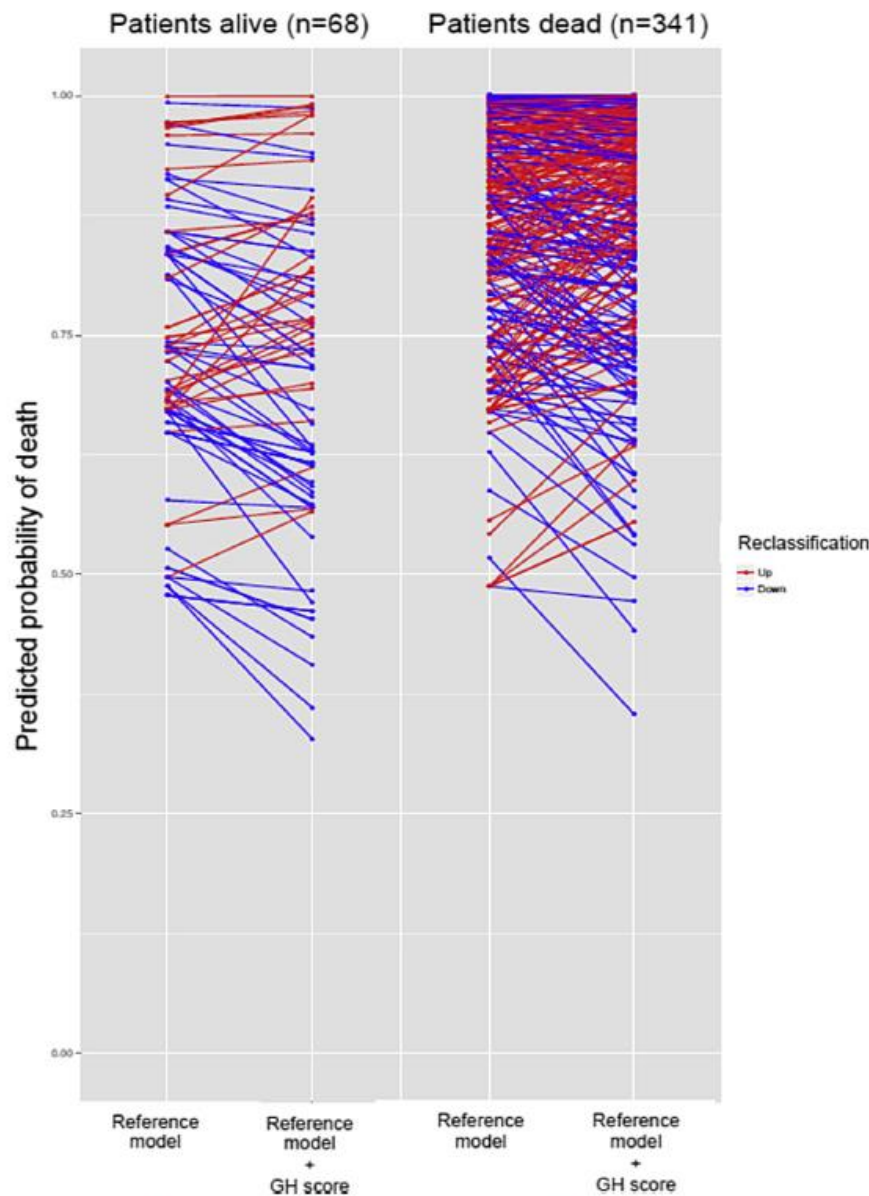


Fig. 1. Additive value of the GH dimension score for reclassifying mortality risk at 24 months post-randomisation (continuous net reclassification improvement). Blue lines in patients who did not die indicate that the GH score had the correct (downward) influence on risk prediction (42/68=61.8%). Conversely, red lines in patients who died indicate a correct (upward) change in risk assessment when using GH score (196/341=57.4%). GH = global health.

proven able to predict treatment efficacy in elderly lung cancer patients, meaning that HRQoL could represent a better tool to identify patients likely to benefit from doublet chemotherapy. However, further research is warranted to validate the questionnaire's predictive value and to define cut-off values.

Our study displayed several limitations. First, the specific lung cancer module QLQ-LC13 questionnaire, which could have improved the HRQoL's prognostic value, was not employed. Secondly, data must be

replicated using external validation study and confirmed in a prospectively recruited cohort. Furthermore, the Instrumental Activities Daily Living index, which explores patients' ability to use public transportation, telephone, drive, etc., was not administered. Finally, our study was not designed to predict the treatment type to be given to the patients, based on HRQoL score.

Our study provides evidence of the additional prognostic value of HRQoL data to identify vulnerable elderly NSCLC subpopulations. The EORTC QLQ-C30

Table 5

Clinical and health-related quality of life dimension scores associated with overall survival using the Q1 and Q3 interquartiles of global health score.

	Number of patients	Number of events	HR (95% CI)	P-value
GH <46	136	123	1.00	<0.0001
46 ≤GH ≤67	228	193	0.42 (0.31–0.59)	
GH >67	87	61	0.67 (0.53–0.85)	
Treatment				
Monochemotherapy	225	199	1.00	
Doublet chemotherapy	226	198	0.64 (0.52–0.79)	<0.0001
Performance status score				
0–1	327	262	0.66 (0.51–0.84)	
2	123	114	1.00	0.0003
Smoking status				
Never smoked	87	68	0.58 (0.44–0.77)	
Ever smoked	264	309	1.00	0.0003
Histology				
Adenocarcinoma	219	178	0.67 (0.50–0.89)	
Squamous	146	131	0.77 (0.57–1.05)	
Other	66	68	1.00	0.025
MMS	441	377	0.97 (0.95–1.01)	0.059
ADL	438	377	0.73 (0.61–0.87)	0.0003

HR = hazard ratio; CI = confidence interval; GH = global health score; MMS = mini-mental state; ADL = activities of daily living.

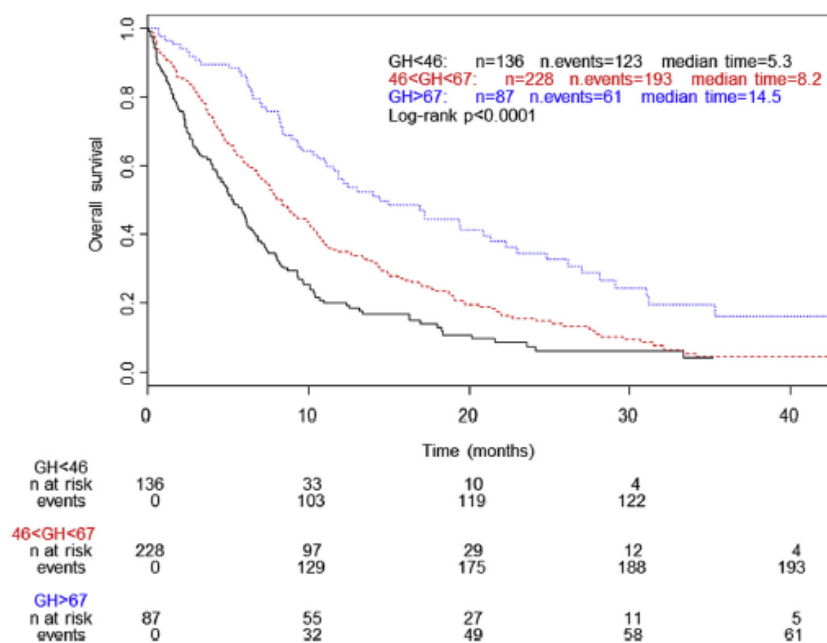


Fig. 2. Kaplan–Meier survival curves according to GH score. GH = global health.

questionnaire could constitute a valuable tool for selecting patients likely to benefit from doublet chemotherapy.

#### Conflict of interest statement

Frédéric Fiteni declares no conflict of interest. Dewi Nemery declares no conflict of interest. Franck Bonnetain reports grants and personal fees from ROCHE, grants, personal fees and non-financial support from NOVARTIS, personal fees from MERCK SERONO,

and personal fees from NESTLE. Fabien Vaylet declares no conflict of interest. Hélène Sennelart declares no conflict of interest. Jean Tredaniel declares no conflict of interest. Denis Moro-Sibilot declares participation to Roche, Eli Lilly, AstraZeneca, Novartis, Pfizer, Boehringer Ingelheim, Amgen, and BMS boards. Dominique Herman declares no conflict of interest. Hélène Laizé declares no conflict of interest. Philippe Masson declares no conflict of interest. Marc Derollez declares no conflict of interest. Christelle Clément-Duchêne declares no conflict of interest. Bernard

**Supplementary Table 1.** Clinical and health-related quality of life dimension scores associated with overall survival using multivariate stepwise backward elimination after univariate analysis

	Number of patients	Number of events	HR (95% CI)	P-value
GH	420	377	0.988 (0.982-0.994)	0.0002
PA	420	377	1.004 (1.00-1.009)	0.046
Treatment				
Monochemotherapy	225	199	1.00	
Doublet chemotherapy	226	198	0.66 (0.53-0.82)	0.0002
Performance status score				
0-1	327	262	0.63 (0.49-0.81)	
2	123	114	1.00	0.0003
Smoking status				
Never smoked	87	68	0.59 (0.44-0.80)	
Ever smoked	264	309	1.00	0.0007
Histology				
Adenocarcinoma	219	178	0.67 (0.49-0.92)	
Squamous	146	131	0.82 (0.59-1.14)	
Other	66	68	1.00	0.037
MMS	441	377	0.97 (0.94-1.00)	0.048
ADL	438	377	0.77 (0.63-0.94)	0.0085

HR = hazard ratio; CI = confidence interval; GH = global health score; PA = pain dimension score; MMS = mini-mental state; ADL = activities of daily living.

**Supplementary Table 2.** Clinical and health-related quality of life dimension scores associated with overall survival using the treatment variable as stratification factor

	Number of patients	Number of events	HR (95% CI)	P-value
GH	420	377	0.986 (0.980-0.992)	<0.0001
Performance status score				
0-1	327	262	0.60 (0.46-0.78)	
2	123	114	1.00	<0.0001
Smoking status				
Never smoked	87	68	0.57 (0.42-0.76)	
Ever smoked	264	309	1.00	0.0002
MMS	441	377	0.96 (0.94-0.99)	0.016
ADL	438	377	0.60 (0.63-0.88)	0.0013

HR = hazard ratio; CI = confidence interval; GH = global health score; MMS = mini-mental state; ADL = activities of daily living.

## *Discussion*

---

« Placez votre main sur un poêle une minute et ça vous semble durer une heure. Asseyez-vous auprès d'une jolie fille une heure et ça vous semble durer une minute. C'est ça la relativité. »

*Albert Einstein*

La SG est le critère de référence mais avec l'augmentation du nombre de traitements efficaces disponibles dans beaucoup de cancers, il est nécessaire d'inclure un nombre plus important de patients et de les suivre plus longtemps ce qui augmente le coût des essais cliniques. Ainsi, les critères de jugement dit intermédiaires centrés sur la tumeur qui sont évalués plus précocement et utilisés comme critères substitutifs de la survie globale sont de plus en plus utilisés mais la plupart d'entre eux ne sont pas validés comme critères de substitution selon la méthodologie de référence. Ainsi la QdV constitue un critère de jugement pour évaluer un bénéfice clinique direct pour le patient. Néanmoins, la QdV est un phénomène complexe et de nombreuses méthodes statistiques ont été proposées mais peu de recommandations ont été données sur les approches statistiques à utiliser selon les situations thérapeutiques.

### **I. Les critères de jugement cliniques dans les essais de phase III de CBNPC avancés**

Un premier objectif a été d'étudier les critères de jugement cliniques dans les essais de phase III de CBNPC avancés.

Nos résultats ont montré qu'une majorité des essais cliniques de 1<sup>ère</sup> ligne de chimiothérapie utilisent la SG comme critère principal (73%) suivie de la SSP (17%). 32% des critères de jugement n'étaient pas du tout définis et seulement 43% était clairement défini. De plus, la progression pour la SSP, le temps jusqu'à progression ou le temps jusqu'à échec du traitement était évaluée le plus fréquemment par les critères RECIST. On peut se demander quelle est la pertinence des critères RECIST avec l'avènement des nouvelles immunothérapies qui n'agirait pas forcément sur la taille tumorale.

La suite de cette première étude sera de proposer une définition consensuelle des critères de jugement dans les essais de phase III de CBNPC avancés à travers le projet DATECAN (Definition for the Assessment of Time-to-event Endpoints in CANcer trials) (7) grâce à une méthode de consensus formalisée.

### **II. Les critères de substitution de la survie globale dans les essais cliniques de cancer pulmonaire**

Nous avons ensuite étudié les critères de substitution de la SG dans les essais cliniques de cancer pulmonaire.



Nous montrons que les techniques utilisées pour valider la substitutivité d'un critère intermédiaire sur la SG sont hétérogènes et peu d'études utilisent la méthode méta-analytique sur données individuelles. Par ailleurs, une seule étude montre que la survie sans maladie et la SSP sont des critères substitutifs de la SG à la phase adjuvante et localement avancée (87). Cette étude nécessite néanmoins confirmation et n'est pas généralisable à de nouvelles thérapeutiques. A la phase métastatique, aucun critère intermédiaire n'a démontré sa substitutivité sur la SG.

La SG reste donc le critère de jugement de référence et la QdV est un critère supplémentaire pour mesurer un réel bénéfice clinique pour le patient.

### **III. Méthodologie d'analyse de la QdV dans les essais de phase III de CBNPC avancés**

Nous avons étudié la méthodologie d'analyse de la QdV dans les essais de phase III de CBNPC avancés. Seulement 49% des essais ont utilisé la QdV comme critère de jugement. De plus la mesure, l'analyse et le report des données de QdV étaient hétérogènes entre les études.

Par exemple, les essais OPTIMAL (34) et lux-lung 3 (33) ont comparé l'efficacité et la tolérance de 2 inhibiteurs de tyrosine kinase (erlotinib et afatinib, respectivement) en comparaison à une chimiothérapie en 1<sup>ère</sup> ligne de traitement chez des patients atteints de CBNPC avancés mutés EGFR. La QdV a été évalué par le questionnaire FACT-L dans l'essai OPTIMAL et par le questionnaire QLQ-C30 associé au module complémentaire LC13 dans l'essai Lux-Lung3. Il est donc difficile de comparer ces 2 essais. Par conséquent, un des challenges serait de promouvoir des recommandations pour la sélection des questionnaires. Par ailleurs, les dimensions cibles, la DMCI et la méthode statistique de prise en compte des données manquantes étaient précisées dans seulement 48,1%, 33,3% et 18,5% respectivement. Concernant les techniques d'analyse, la plus utilisée est le changement moyen par rapport à baseline qui n'est pas une méthode longitudinale et ne devrait pas être utilisée. De plus, lorsque les 2 techniques standards sont utilisées : les modèles mixtes d'analyse de variance sur mesures répétées et le TJD, nous avons constaté la faiblesse et l'hétérogénéité de leurs définitions.

Des PROs CONSORT (88) ont été publiés mais ils ne traitent pas ou peu de l'analyse statistique de la QdV. Par conséquent, des recommandations sur l'analyse statistique de la QdV devraient être développées. Quelques recommandations ont été données sur les

définitions de TJD à appliquer selon les situations thérapeutiques par Anota et al. (77). Ainsi, en situation adjuvante, le TJD simple en tant qu'état transitoire semble la définition la plus adaptée puisque le patient a de grandes chances de guérir de son cancer. En revanche, en situation avancée ou métastatique, il paraît plus pertinent d'étudier le TJD définitif, intégrant ou non le décès dans la définition de l'évènement, selon la définition de Bonnetain *et al* (89).

Un projet a été initié par l'EORTC afin de proposer une standardisation de l'analyse de la QdV. Cette standardisation facilitera la comparaison entre les essais cliniques et permettra à long terme une meilleure considération des résultats de QdV.

#### **IV. La qualité de vie relative à la santé chez les patients âgés atteints d'un carcinome bronchique non à petites cellules avancé traité par le doublet de chimiothérapie carboplatine et paclitaxel hebdomadaire en comparaison à une monochimiothérapie**

Nous avons comparé longitudinalement la QdV chez 451 patients âgés de 70-89 ans atteints de CBNPC avancés randomisés entre 4 cycles de carboplatin plus paclitaxel hebdomadaire et 5 cycles de vinorelbine ou gemcitabine selon la technique du temps jusqu'à détérioration définitive d'un score de QdV. Le TJD a été défini par le temps entre la randomisation et la première décroissance d'un score du QLQ-C30  $\geq 5$  points sans amélioration ultérieure d'un score du QLQ-C30  $\geq 5$  points ou autre donnée de QdV disponible. Nous avons montré que malgré une augmentation de la toxicité, le doublet de chimiothérapie ne diminue pas le TJD chez les patients âgés atteints de CBNPC avancés. De plus, l'augmentation de la SG pour le doublet de chimiothérapie est associée à une augmentation du TJD pour 2 dimensions de QdV, à savoir le fonctionnement physique et les nausées/vomissements.

Cette étude a des implications cliniques et méthodologiques :

- Clinique : le bénéfice du doublet de chimiothérapie pour les patients est renforcé par les données de QdV ;
- Méthodologiques : le TJD est une technique d'analyse faisable dans les essais de phase III de CBNPC et fournit des résultats faciles d'interprétation ;

Dans notre étude l'évènement « décès » n'était pas pris en compte dans la définition du TJD afin d'éviter le risque compétitif avec l'évènement « détérioration ». Cette définition nécessite d'être encore explorée afin de standardiser l'analyse longitudinale selon la méthode du TJD dans les CBNPC avancés. Une autre cohorte issue de l'essai IFCT-1201 pourrait être analysée pour valider cette définition.

## **V. La qualité de vie relative à la santé comme co-critère de jugement principal dans les essais cliniques en oncologie**

Dans notre travail nous avons discuté les problèmes méthodologiques soulevés par un design associant un critère de jugement intermédiaire (par exemple la PFS) et la QdV. Le choix du nombre et du type de dimensions de QdV ainsi que les règles décisionnelles dans un essai avec un tel design sont des questions majeures. On pourrait décider que l'essai est positif si le critère intermédiaire et au moins une dimension de QdV (sans détérioration des autres dimensions) sont positifs. Une autre approche serait d'avoir le critère intermédiaire et toutes les dimensions de QdV positifs. Par ailleurs, il est important de ne pas perdre l'aspect multidimensionnel de la QdV. De ce fait, l'inclusion de plus d'une dimension avec au moins une dimension de QdV globale doit être envisagée systématiquement. Tout ceci doit être discuté au préalable entre les méthodologistes et les cliniciens et écrit de manière précise dans le protocole.

Le contrôle de l'erreur de type I est aussi un problème essentiel. Dans le cas où tous les critères et dimensions de QdV doivent être positifs, toutes les hypothèses nulles doivent être rejetées et aucun ajustement est nécessaire, il y a cependant une perte de puissance qui doit être prise en compte dans le calcul du nombre de sujets nécessaires. Dans le cas où au moins un critère doit être positif, de multiples procédures ont été développées : la méthode de Bonferroni avec une répartition équivalente du risque entre les critères, la méthode Bonferroni-Holm avec une possibilité de pondérer le risque en fonction de l'importance clinique des critères.

Enfin, le calcul du nombre de sujets nécessaires doit prendre en compte l'inflation du risque de type I. Une approche possible serait de calculer le nombre de sujet nécessaire pour chaque critère et choisir le nombre le plus important afin de répondre à toutes les questions d'intérêt.

## **VI. Valeur pronostique de la qualité de vie relative à la santé sur la survie globale chez les patients âgés atteints de CBNPC**

Dans notre travail, nous avons démontré la valeur pronostique de la dimension santé globale de QdV à baseline du questionnaire EORTC QLQ-C30 sur la SG chez les patients âgés atteints de CBNPC avancés. Ce résultat est en cohérence avec les études de Sloan et al.

(85) et Jacot et al. (86) qui ont montré la valeur pronostique de la QdV sur la SG au diagnostic de cancer du poumon. Cependant, ces 2 études n'avaient pas pris en compte l'âge et le stade de la maladie et n'utilisaient pas le questionnaire EORTC QLQ-C30.

Dans notre modèle les scores MMS et ADL sont significativement associés à la SG. Il est intéressant de remarquer que la dimension de fonctionnement physique est significativement associée à la SG lorsque l'on retire le score ADL du modèle. Par conséquent, on pourrait imaginer que l'évaluation gériatrique d'un patient atteint d'un CBNPC avancé au début de la prise en charge serait effectuée à l'aide du seul questionnaire EORTC QLQ-C30 afin de limiter le nombre de questionnaires.

De plus, nous avons étudié la valeur prédictive du score de santé globale sur l'efficacité de la bichimiothérapie. Nous avons découpé le score de santé globale à baseline en quartiles et l'analyse en sous-groupe suggère que les patients qui ont un score inférieur à 46 à baseline ne bénéficieraient pas de la bichimiothérapie. La mesure de la QdV à baseline pourrait être un outil prédictif de l'efficacité du traitement et pourrait aider les cliniciens dans leur prise de décision.

L'ensemble de ces résultats nécessitent une validation par une cohorte externe. Ceci pourrait faire l'objet d'un travail ultérieur à partir d'essais cliniques conduits chez le sujet âgé atteint d'un CBNPC avancé comme par exemple l'essai IFCT-1201 MODEL.

## *Conclusion*

---

« L'ennui dans ce monde, c'est que les idiots sont sûr d'eux et les gens sensés pleins de doutes. »

**Bertrand Russel**

Ce travail a porté sur l'étude des critères de jugement centrés sur le patient dans les essais cliniques en oncologie thoracique et plus particulièrement l'analyse de la QdV relative à la santé.

Les différents travaux menés ont montré la nécessité d'obtenir une définition consensuelle des critères de jugement et qui fera l'objet du projet « DATECAN poumon ».

Par ailleurs, nous avons montré l'absence de substitutivité des critères intermédiaires centrés sur la tumeur sur la SG et donc l'intérêt de la mesure la QdV afin d'évaluer le réel bénéfice clinique pour le patient.

Cependant, il existe une hétérogénéité de la mesure, l'analyse statistique et le report des données de QdV entre les études. Des PROs CONSORT ont été publiés mais ils traitent peu de l'analyse statistique. Dans une optique de standardisation, un projet a été initié par l'EORTC dans le but de proposer une standardisation de l'analyse de la QdV.

Le TJD est de plus en plus souvent utilisé en tant que méthode d'analyse longitudinale de la QdV dans les essais de phase III en cancérologie. Cette méthode offre l'avantage de présenter des résultats qui sont familiers pour les cliniciens, résumés sous forme d'Hazard Ratio et de médianes de survie. Cette technique a été utilisée pour comparer longitudinalement la QdV chez 451 patients âgés de 70-89 ans atteints de CBNPC avancés randomisés entre 4 cycles de carboplatin plus paclitaxel hebdomadaire et 5 cycles de vinorelbine ou gemcitabine. D'un point de vue clinique, le bénéfice du doublet de chimiothérapie pour les patients est renforcé par les données de QdV et d'un point de vue méthodologique, le TJD est une technique d'analyse faisable dans les essais de phase III de CBNPC et qui fournit des résultats faciles d'interprétation.

Par ailleurs, une nouvelle approche proposée dans notre travail est l'utilisation des co-critères de jugement associant un critère centré sur la tumeur et la QdV. Un tel design pose de nombreux problèmes méthodologiques notamment les règles décisionnelles et le calcul du nombre de sujets nécessaires. Il est donc indispensable de poursuivre la réflexion méthodologique afin d'améliorer et standardiser la modélisation des essais utilisant ce genre d'approche.

Enfin, la mesure de la QdV à baseline dans les essais de phase III en oncologie thoracique a probablement une valeur pronostique sur la SG. De plus, elle aurait une valeur prédictive sur l'efficacité de la chimiothérapie en termes de SG et pourrait aider les cliniciens dans leur prise de décision thérapeutique.

## Références

---

1. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med*. 20 oct 2005;353(16):1673-84.
2. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med*. 15 mars 2001;344(11):783-92.
3. Gianni L, Dafni U, Gelber RD, Azambuja E, Muehlbauer S, Goldhirsch A, et al. Treatment with trastuzumab for 1 year after adjuvant chemotherapy in patients with HER2-positive early breast cancer: a 4-year follow-up of a randomised controlled trial. *Lancet Oncol*. mars 2011;12(3):236-44.
4. Joensuu H, Bono P, Kataja V, Alanko T, Kokko R, Asola R, et al. Fluorouracil, epirubicin, and cyclophosphamide with either docetaxel or vinorelbine, with or without trastuzumab, as adjuvant treatments of breast cancer: final results of the FinHer Trial. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 déc 2009;27(34):5685-92.
5. Birgisson H, Wallin U, Holmberg L, Glimelius B. Survival endpoints in colorectal cancer and the effect of second primary other cancer on disease free survival. *BMC Cancer*. 2011;11:438.
6. Van Cutsem E, Labianca R, Bodoky G, Barone C, Aranda E, Nordlinger B, et al. Randomized phase III trial comparing biweekly infusional fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 juill 2009;27(19):3117-25.
7. Bellera CA, Pulido M, Gourgou S, Collette L, Doussau A, Kramar A, et al. 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 Oxf Engl 1990*. mars 2013;49(4):769-81.
8. Bhattacharya S, Fyfe G, Gray RJ, Sargent DJ. Role of sensitivity analyses in assessing progression-free survival in late-stage oncology trials. *J Clin Oncol Off J Am Soc Clin Oncol*. 10 déc 2009;27(35):5958-64.
9. Panageas KS, Ben-Porat L, Dickler MN, Chapman PB, Schrag D. When you look matters: the effect of assessment schedule on progression-free survival. *J Natl Cancer Inst*. 21 mars 2007;99(6):428-32.
10. Ciani O, Davis S, Tappenden P, Garside R, Stein K, Cantrell A, et al. Validation of surrogate endpoints in advanced solid tumors: systematic review of statistical methods, results, and implications for policy makers. *Int J Technol Assess Health Care*. juill 2014;30(3):312-24.



11. Sherrill B, Kaye JA, Sandin R, Cappelleri JC, Chen C. Review of meta-analyses evaluating surrogate endpoints for overall survival in oncology. *OncoTargets Ther.* 2012;5:287-96.
12. Miller K, Wang M, Gralow J, Dickler M, Cobleigh M, Perez EA, et al. Paclitaxel plus bevacizumab versus paclitaxel alone for metastatic breast cancer. *N Engl J Med.* 27 déc 2007;357(26):2666-76.
13. Miles DW. Bevacizumab in breast cancer: fundamental questions remain. *Lancet Oncol.* févr 2013;14(2):99-101.
14. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med.* avr 1989;8(4):431-40.
15. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Stat Med.* 30 janv 1992;11(2):167-78.
16. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther.* mars 2001;69(3):89-95.
17. Ohorodnyk P, Eisenhauer EA, Booth CM. Clinical benefit in oncology trials: is this a patient-centred or tumour-centred end-point? *Eur J Cancer Oxf Engl* 1990. sept 2009;45(13):2249-52.
18. Booth CM, Ohorodnyk P, Eisenhauer EA. Call for clarity in the reporting of benefit associated with anticancer therapies. *J Clin Oncol Off J Am Soc Clin Oncol.* 20 nov 2009;27(33):e213-4.
19. Clinical Trial Endpoints for the Approval of Non- Small Cell Lung Cancer Drugs and Biologics - UCM259421.pdf [Internet]. [cité 3 juin 2015]. Disponible sur: <http://www.fda.gov/downloads/Drugs/Guidances/UCM259421.pdf>
20. Saad ED, Buyse M. Overall survival: patient outcome, therapeutic objective, clinical trial end point, or public health measure? *J Clin Oncol Off J Am Soc Clin Oncol.* 20 mai 2012;30(15):1750-4.
21. Doward LC, McKenna SP. Defining patient-reported outcomes. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* oct 2004;7 Suppl 1:S4-8.
22. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA.* 4 janv 1995;273(1):59-65.
23. Osoba D. Translating the science of patient-reported outcomes assessment into clinical practice. *J Natl Cancer Inst Monogr.* 2007;(37):5-11.
24. Peppercorn JM, Smith TJ, Helft PR, Debono DJ, Berry SR, Wollins DS, et al. American society of clinical oncology statement: toward individualized care for patients with advanced cancer. *J Clin Oncol Off J Am Soc Clin Oncol.* 20 févr 2011;29(6):755-60.

25. Mathoulin-Pelissier S, Gourgou-Bourgade S, Bonnetain F, Kramar A. Survival end point reporting in randomized cancer clinical trials: a review of major journals. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 août 2008;26(22):3721-6.
26. Yardley DA, Noguchi S, Pritchard KI, Burris HA, Baselga J, Gnant M, et al. Everolimus plus exemestane in postmenopausal patients with HR(+) breast cancer: BOLERO-2 final progression-free survival analysis. *Adv Ther*. oct 2013;30(10):870-84.
27. Burris HA 3rd, Lebrun F, Rugo HS, Beck JT, Piccart M, Neven P, et al. Health-related quality of life of patients with advanced breast cancer treated with everolimus plus exemestane versus placebo plus exemestane in the phase 3, randomized, controlled, BOLERO-2 trial. *Cancer*. 15 mai 2013;119(10):1908-15.
28. Verma S, Miles D, Gianni L, Krop IE, Welslau M, Baselga J, et al. Trastuzumab emtansine for HER2-positive advanced breast cancer. *N Engl J Med*. 8 nov 2012;367(19):1783-91.
29. Welslau M, Diéras V, Sohn J-H, Hurvitz SA, Lalla D, Fang L, et al. Patient-reported outcomes from EMILIA, a randomized phase 3 study of trastuzumab emtansine (T-DM1) versus capecitabine and lapatinib in human epidermal growth factor receptor 2-positive locally advanced or metastatic breast cancer. *Cancer*. 1 mars 2014;120(5):642-51.
30. Baselga J, Cortés J, Kim S-B, Im S-A, Hegg R, Im Y-H, et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. *N Engl J Med*. 12 janv 2012;366(2):109-19.
31. Cortés J, Baselga J, Im Y-H, Im S-A, Pivot X, Ross G, et al. Health-related quality-of-life assessment in CLEOPATRA, a phase III study combining pertuzumab with trastuzumab and docetaxel in metastatic breast cancer. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. oct 2013;24(10):2630-5.
32. Wu Y-L, Zhou C, Hu C-P, Feng J, Lu S, Huang Y, et al. Afatinib versus cisplatin plus gemcitabine for first-line treatment of Asian patients with advanced non-small-cell lung cancer harbouring EGFR mutations (LUX-Lung 6): an open-label, randomised phase 3 trial. *Lancet Oncol*. févr 2014;15(2):213-22.
33. Yang JC-H, Hirsh V, Schuler M, Yamamoto N, O'Byrne KJ, Mok TSK, et al. Symptom control and quality of life in LUX-Lung 3: a phase III study of afatinib or cisplatin/pemetrexed in patients with advanced lung adenocarcinoma with EGFR mutations. *J Clin Oncol Off J Am Soc Clin Oncol*. 20 sept 2013;31(27):3342-50.
34. Chen G, Feng J, Zhou C, Wu Y-L, Liu X-Q, Wang C, et al. Quality of life (QoL) analyses from OPTIMAL (CTONG-0802), a phase III, randomised, open-label study of first-line erlotinib versus chemotherapy in patients with advanced EGFR mutation-positive non-small-cell lung cancer (NSCLC). *Ann Oncol Off J Eur Soc Med Oncol ESMO*. juin 2013;24(6):1615-22.

35. Zhou C, Wu Y-L, Chen G, Feng J, Liu X-Q, Wang C, et al. Erlotinib versus chemotherapy as first-line treatment for patients with advanced EGFR mutation-positive non-small-cell lung cancer (OPTIMAL, CTONG-0802): a multicentre, open-label, randomised, phase 3 study. *Lancet Oncol*. août 2011;12(8):735-42.
36. Barlesi F, Scherpereel A, Rittmeyer A, Pazzola A, Ferrer Tur N, Kim J-H, et al. Randomized phase III trial of maintenance bevacizumab with or without pemetrexed after first-line induction with bevacizumab, cisplatin, and pemetrexed in advanced nonsquamous non-small-cell lung cancer: AVAPERL (MO22089). *J Clin Oncol Off J Am Soc Clin Oncol*. 20 août 2013;31(24):3004-11.
37. Rittmeyer A, Gorbunova V, Vikström A, Scherpereel A, Kim J-H, Ahn M-J, et al. Health-related quality of life in patients with advanced nonsquamous non-small-cell lung cancer receiving bevacizumab or bevacizumab-plus-pemetrexed maintenance therapy in AVAPERL (MO22089). *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. nov 2013;8(11):1409-16.
38. Pujade-Lauraine E, Hilpert F, Weber B, Reuss A, Poveda A, Kristensen G, et al. Bevacizumab combined with chemotherapy for platinum-resistant recurrent ovarian cancer: The AURELIA open-label randomized phase III trial. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 mai 2014;32(13):1302-8.
39. Stockler MR, Hilpert F, Friedlander M, King MT, Wenzel L, Lee CK, et al. Patient-reported outcome results from the open-label phase III AURELIA trial evaluating bevacizumab-containing therapy for platinum-resistant ovarian cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 mai 2014;32(13):1309-16.
40. Burger RA, Brady MF, Bookman MA, Fleming GF, Monk BJ, Huang H, et al. Incorporation of bevacizumab in the primary treatment of ovarian cancer. *N Engl J Med*. 29 déc 2011;365(26):2473-83.
41. Djärv T, Metcalfe C, Avery KNL, Lagergren P, Blazeby JM. Prognostic value of changes in health-related quality of life scores during curative treatment for esophagogastric cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 avr 2010;28(10):1666-70.
42. Conroy T, Desseigne F, Ychou M, Bouché O, Guimbaud R, Bécouarn Y, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. *N Engl J Med*. 12 mai 2011;364(19):1817-25.
43. Gourgou-Bourgade S, Bascoul-Mollevi C, Desseigne F, Ychou M, Bouché O, Guimbaud R, 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*. 1 janv 2013;31(1):23-9.
44. Hussain M, Tangen CM, Berry DL, Higano CS, Crawford ED, Liu G, et al. Intermittent versus continuous androgen deprivation in prostate cancer. *N Engl J Med*. 4 avr 2013;368(14):1314-25.

45. von Plessen C, Bergman B, Andresen O, Bremnes RM, Sundstrom S, Gilleryd M, et al. Palliative chemotherapy beyond three courses conveys no survival or consistent quality-of-life benefits in advanced non-small-cell lung cancer. *Br J Cancer*. 23 oct 2006;95(8):966-73.
46. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, 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*. 3 mars 1993;85(5):365-76.
47. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. *J Clin Oncol Off J Am Soc Clin Oncol*. mars 1993;11(3):570-9.
48. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. *Med Care*. mars 1989;27(3 Suppl):S178-89.
49. Katz JN, Larson MG, Phillips CB, Fossel AH, Liang MH. Comparative measurement sensitivity of short and longer health status instruments. *Med Care*. oct 1992;30(10):917-25.
50. Buskirk TD, Stein KD. Telephone vs. mail survey gives different SF-36 quality-of-life scores among cancer survivors. *J Clin Epidemiol*. oct 2008;61(10):1049-55.
51. Weinberger M, Oddone EZ, Samsa GP, Landsman PB. Are health-related quality-of-life measures affected by the mode of administration? *J Clin Epidemiol*. févr 1996;49(2):135-40.
52. Holzner B, Giesinger JM, Pinggera J, Zugal S, Schöpf F, Oberguggenberger AS, et al. The Computer-based Health Evaluation Software (CHES): a software for electronic patient-reported outcome monitoring. *BMC Med Inform Decis Mak*. 2012;12:126.
53. Millsap RE. Comments on methods for the investigation of measurement bias in the Mini-Mental State Examination. *Med Care*. nov 2006;44(11 Suppl 3):S171-5.
54. Noll RB, Fairclough D. Health-related quality of life: developmental and psychometric issues. *J Pediatr*. juill 2004;145(1):8-9.
55. Wiklund I. Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life. *Fundam Clin Pharmacol*. juin 2004;18(3):351-63.
56. Brooks MM, Jenkins LS, Schron EB, Steinberg JS, Cross JA, Paeth DS. Quality of life at baseline: is assessment after randomization valid? The AVID Investigators. The Antiarrhythmics Versus Implantable Defibrillators. *Med Care*. oct 1998;36(10):1515-9.
57. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. *J Clin Epidemiol*. mai 2010;63(5):524-34.

58. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. *J Clin Epidemiol.* mai 2003;56(5):395-407.
59. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* févr 2008;61(2):102-9.
60. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials.* déc 1989;10(4):407-15.
61. Yost KJ, Sorensen MV, Hahn EA, Glendenning GA, Gnanasakthy A, Cella D. Using multiple anchor- and distribution-based estimates to evaluate clinically meaningful change on the Functional Assessment of Cancer Therapy-Biologic Response Modifiers (FACT-BRM) instrument. *Value Health J Int Soc Pharmacoeconomics Outcomes Res.* avr 2005;8(2):117-27.
62. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care.* mai 2003;41(5):582-92.
63. Wyrwich KW, Nienaber NA, Tierney WM, Wolinsky FD. Linking clinical relevance and statistical significance in evaluating intra-individual changes in health-related quality of life. *Med Care.* mai 1999;37(5):469-78.
64. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an SEM-based criterion for identifying meaningful intra-individual changes in health-related quality of life. *J Clin Epidemiol.* sept 1999;52(9):861-73.
65. 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.* janv 1998;16(1):139-44.
66. Cocks K, King MT, Velikova G, Martyn St-James M, Fayers PM, Brown JM. Evidence-based guidelines for determination of sample size and interpretation of the European Organisation for the Research and Treatment of Cancer Quality of Life Questionnaire Core 30. *J Clin Oncol Off J Am Soc Clin Oncol.* 1 janv 2011;29(1):89-96.
67. Eton DT, Fairclough DL, Cella D, Yount SE, Bonomi P, Johnson DH, et al. Early change in patient-reported health during lung cancer chemotherapy predicts clinical outcomes beyond those predicted by baseline report: results from Eastern Cooperative Oncology Group Study 5592. *J Clin Oncol Off J Am Soc Clin Oncol.* 15 avr 2003;21(8):1536-43.
68. Cella D, Eton DT, Fairclough DL, Bonomi P, Heyes AE, Silberman C, et al. What is a clinically meaningful change on the Functional Assessment of Cancer Therapy-Lung (FACT-L) Questionnaire? Results from Eastern Cooperative Oncology Group (ECOG) Study 5592. *J Clin Epidemiol.* mars 2002;55(3):285-95.

69. Maringwa JT, Quinten C, King M, Ringash J, Osoba D, Coens C, et al. Minimal important differences for interpreting health-related quality of life scores from the EORTC QLQ-C30 in lung cancer patients participating in randomized controlled trials. *Support Care Cancer Off J Multinatl Assoc Support Care Cancer*. nov 2011;19(11):1753-60.
70. Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality of life scores: differences between improvement and worsening. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil*. mai 2002;11(3):207-21.
71. Hong F, Bosco JLF, Bush N, Berry DL. Patient self-appraisal of change and minimal clinically important difference on the European organization for the research and treatment of cancer quality of life questionnaire core 30 before and during cancer therapy. *BMC Cancer*. 2013;13:165.
72. Ward MM, Guthrie LC, Alba M. Dependence of the minimal clinically important improvement on the baseline value is a consequence of floor and ceiling effects and not different expectations by patients. *J Clin Epidemiol*. juin 2014;67(6):689-96.
73. Bonnetain F. [Health related quality of life and endpoints in oncology]. *Cancer Radiothérapie J Société Fr Radiothérapie Oncol*. oct 2010;14(6-7):515-8.
74. Kabbinavar FF, Wallace JF, Holmgren E, Yi J, Cella D, Yost KJ, et al. Health-related quality of life impact of bevacizumab when combined with irinotecan, 5-fluorouracil, and leucovorin or 5-fluorouracil and leucovorin for metastatic colorectal cancer. *The oncologist*. sept 2008;13(9):1021-9.
75. Awad L, Zuber E, Mesbah M. Applying Survival Data Methodology to Analyze Longitudinal Quality of Life Data. In: Mesbah M, Cole BF, Lee M-LT, éditeurs. *Statistical Methods for Quality of Life Studies: Design, Measurements and Analysis* [Internet]. Boston, MA: Springer US; 2002. p. 231-43. Disponible sur: [http://dx.doi.org/10.1007/978-1-4757-3625-0\\_18](http://dx.doi.org/10.1007/978-1-4757-3625-0_18)
76. Hamidou Z, Dabakuyo TS, Mercier M, Fraisse J, Causeret S, Tixier H, et al. Time to deterioration in quality of life score as a modality of longitudinal analysis in patients with breast cancer. *The oncologist*. 2011;16(10):1458-68.
77. Anota A, Hamidou Z, Paget-Bailly S, Chibaudel B, Bascoul-Mollevis C, Auquier P, 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 Int J Qual Life Asp Treat Care Rehabil*. 26 nov 2013;
78. Wisløff F, Hjorth M. Health-related quality of life assessed before and during chemotherapy predicts for survival in multiple myeloma. *Nordic Myeloma Study Group*. *Br J Haematol*. avr 1997;97(1):29-37.
79. Efficace F, Bottomley A, Smit EF, Lianes P, Legrand C, Debruyne C, et al. Is a patient's self-reported health-related quality of life a prognostic factor for survival in

- non-small-cell lung cancer patients? A multivariate analysis of prognostic factors of EORTC study 08975. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. nov 2006;17(11):1698-704.
80. Montazeri A. Quality of life data as prognostic indicators of survival in cancer patients: an overview of the literature from 1982 to 2008. *Health Qual Life Outcomes*. 2009;7:102.
  81. Roychowdhury DF, Hayden A, Liepa AM. Health-related quality-of-life parameters as independent prognostic factors in advanced or metastatic bladder cancer. *J Clin Oncol Off J Am Soc Clin Oncol*. 15 févr 2003;21(4):673-8.
  82. Luoma M-L, Hakamies-Blomqvist L, Sjöström J, Pluzanska A, Ottoson S, Mouridsen H, et al. Prognostic value of quality of life scores for time to progression (TTP) and overall survival time (OS) in advanced breast cancer. *Eur J Cancer Oxf Engl* 1990. juill 2003;39(10):1370-6.
  83. Cella D, Bushmakina AG, Cappelleri JC, Charbonneau C, Michaelson MD, Motzer RJ. Baseline quality of life as a prognostic survival tool in patients receiving sunitinib for metastatic renal cell carcinoma. *Br J Cancer*. 14 févr 2012;106(4):646-50.
  84. Movsas B, Moughan J, Sarna L, Langer C, Werner-Wasik M, Nicolaou N, et al. Quality of life supersedes the classic prognosticators for long-term survival in locally advanced non-small-cell lung cancer: an analysis of RTOG 9801. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 déc 2009;27(34):5816-22.
  85. Sloan JA, Zhao X, Novotny PJ, Wampfler J, Garces Y, Clark MM, et al. Relationship between deficits in overall quality of life and non-small-cell lung cancer survival. *J Clin Oncol Off J Am Soc Clin Oncol*. 1 mai 2012;30(13):1498-504.
  86. Jacot W, Colinet B, Bertrand D, Lacombe S, Bozonnat M-C, Daurès J-P, et al. Quality of life and comorbidity score as prognostic determinants in non-small-cell lung cancer patients. *Ann Oncol Off J Eur Soc Med Oncol ESMO*. août 2008;19(8):1458-64.
  87. Mauguén A, Pignon J-P, Burdett S, Domerg C, Fisher D, Paulus R, et al. Surrogate endpoints for overall survival in chemotherapy and radiotherapy trials in operable and locally advanced lung cancer: a re-analysis of meta-analyses of individual patients' data. *Lancet Oncol*. juin 2013;14(7):619-26.
  88. Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD, et al. Reporting of patient-reported outcomes in randomized trials: the CONSORT PRO extension. *JAMA J Am Med Assoc*. 27 févr 2013;309(8):814-22.
  89. Bonnetain F, Dahan L, Maillard E, Ychou M, Mitry E, Hammel P, 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 Oxf Engl* 1990. oct 2010;46(15):2753-62.