L'analyse de texture : un nouveau biomarqueur en oncologie?

Caroline Caramella

université PARIS-SACLAY

L'analyse de texture : un nouveau biomarqueur en oncologie ?

Texture analysis : a new biomarker in oncology ?

Thèse de doctorat de l'université Paris-Saclay

École doctorale n° 582 cancérologie : biologie - médecine - santé (CBMS)
Spécialité de doctorat: recherche clinique, innovation technologique, santé publique
Unité de recherche : Université Paris-Saclay, CEA, CNRS, Inserm, Laboratoire d'Imagerie Biomédicale Multimodale Paris Saclay, 91401, Orsay, France
Référent : Faculté de médecine

Thèse présentée et soutenue à Paris-Saclay, le 17/12/2021, par

Caroline CARAMELLA

Composition du Jury

Mathieu LEDERLIN PUPH, CHU Rennes	Président du jury
Benjamin BESSE PUPH, Université Paris Saclay	Membre invité
Marie WISLEZ PUPH, Hôpital Cochin, APHP	Rapporteur & Examinatrice
Alain LUCIANI PUPH, Hôpital H. Mondor, APHP	Rapporteur & Examinateur
Elie FADEL PUPH, Hôpital M. Lannelongue, GHPSJ, Université Paris Saclay	Examinateur

Direction de la thèse

Nathalie LASSAU PUPH, Université Paris Saclay	Directrice de thèse
Stéphanie PITRE-CHAMPAGNAT PhD, HDR, Université Paris Saclay	Co-directrice de thèse

Thèse de doctorat

NNT : 2021UPASL119

Titre : L'analyse de texture : un nouveau biomarqueur en oncologie ?

Mots clés : imagerie, oncologie, biomarqueurs, radiomique, immunothérapie

Ce travail de thèse s'inscrit initialement dans le domaine de l'imagerie tomodensitométrique quantitative. En effet, la numérisation des examens d'imagerie permet d'extraire des images des données chiffrées, ou indices de texture, qui pourraient caractériser plus finement les tissus, en particulier tumoraux, afin d'en extraire des paramètres d'agressivité ou des biomarqueurs prédictifs de réponse aux traitements. Néanmoins, l'obtention des images médicales en tomodensitométrie n'est pas homogène et fait intervenir de multiples paramètres techniques liés aux machines scanographiques et qui pourraient être à l'origine de variation des indices de texture inter et intra-scanner. Mon travail a donc consisté à l'identification de ces sources de variation pour proposer, à terme, une méthode de calibration robuste et standardisée en TDM.

En parallèle des avancées dans ce domaine appelé la radiomique, de nouveaux traitements anticancéreux, faisant appel non plus à la destruction directe des cellules tumorales mais à la modification des interactions entre cellules tumorales et cellules immunitaires (ou immunothérapie) ont été découverts. Il s'agit d'une révolution dans le traitement des patients atteints de maladies métastatiques. L'évaluation de la réponse à ce nouveau type de traitement a été rapidement compliquée par la constatation de réponses atypiques : pseudoprogression (augmentation initiale des lésions suivies d'une stabilisation ou d'une réponse objective) et hyperprogression (accélération de la cinétique tumorale au début du traitement). Dans le cadre de ma thèse, je me suis donc également intéressée à ces nouveaux challenges de l'évaluation tumorale.

Enfin, la découverte d'anomalies génomiques au sein des cellules tumorales dans les adénocarcinomes bronchiques a également été à l'origine du développement de nouvelles stratégies thérapeutiques (médecine personnalisée). Néanmoins, la recherche de ce type d'anomalie nécessite du matériel tumoral de qualité et en quantité suffisantes, et environ 20% des patients ne peuvent pas bénéficier de cette approche. La corrélation phénotype / génotype pourrait permettre d'affiner les recherches d'anomalies moléculaires chez les patients qui ont une présentation de la maladie évoquant la présence d'une anomalie moléculaire. Mon travail a donc consisté en une intégration des données cliniques et radiologiques et au développement d'un algorithme d'intelligence artificielle pour permettre de prédire la probabilité de trouver une anomalie génomique ciblable.

Title : Texture analysis : a new biomarker in oncology ?

Keywords : imaging, oncology, biomarkers, radiomics,, immunotherapy

Abstract This thesis initially addresses the field of quantitative CT imaging. The digitization of imaging examinations alllows to extract quantified data, or texture indices, from images, which could characterize more precisely the tissues, in particular tumors, in order to extract parameters of aggressiveness or predictive biomarkers. Nevertheless, the obtaining of medical images in CT is not homogeneous and involves multiple technical parameters related to the scanning machines and which could be at the origin of inter and intra-scanner texture index variations. My work therefore consisted in identifying these sources of variation in order to propose a robust and standardized calibration method for CT.

In parallel with advances in this field called radiomics, new cancer treatments have been discovered that no longer involve the direct destruction of tumor cells but rather the modification of interactions between tumor cells and immune cells (or immunotherapy). This is a revolution in the treatment of patients with metastatic disease.

The evaluation of the response to this new type of treatment was quickly complicated by the observation of atypical responses: pseudoprogression (initial increase in lesions followed by stabilization or objective response) and hyperprogression (acceleration of tumor kinetics at the beginning of the treatment). In the framework of my thesis, I was therefore also interested in these new challenges of tumor evaluation.

Finally, the discovery of genomic abnormalities within tumor cells in lung adenocarcinoma has also led to the development of new therapeutic strategies (personalized medicine). Nevertheless, the search for this type of abnormality requires tumor material of sufficient quality and quantity, and about 20% of patients cannot benefit from this approach. The phenotype/genotype correlation could allow to refine the search for molecular abnormalities in patients who have a presentation of the disease evoking the presence of a molecular abnormality. My work therefore consisted of integrating clinical and radiological data and developing an artificial intelligence algorithm to predict the probability of finding a targetable genomic anomaly.

Table des matières

Chapitre 1: La quantification des images radiologiques : les biomarqueurs en oncologie

Introduction

Rapidement après la découverte des rayons X par Wilhelm Röntgen en 1895, les premiers « services de radiologie » ouvrent dans le monde. Le déclenchement de la guerre de 14-18 va également appuyer l'intérêt de disposer sur le champ de bataille via des ambulances de radiologie, développées entre autre par Marie Curie, d'une imagerie rapide des corps étrangers métalliques afin d'en guider l'extraction. Après-guerre, les médecins qui ont eu accès à cette technique contribuent à l'installation d'appareils dans les hôpitaux et à la formation du personnel, radiologues et manipulateurs. Les progrès techniques constants qui améliorent les images et l'invention du produit de contraste iodé dans les années 20 élargissent considérablement les indications de l'utilisation des rayons X en médecine. Dans les années 60, la tomodensitométrie est développée, et Hounsfield invente l'échelle de densité qui porte son nom (de -1000 à +1000) ; les premières images du cerveau sont faites en 1973 ; les perfectionnements des techniques et la conjonction avec les progrès informatiques permettent dans les années 90 la numérisation des images.

L'imagerie médicale est un domaine en transformation perpétuelle et à la croisée des applications médicales et technologiques. La rapidité, la disponibilité et la précision de l'obtention d'images du corps humain sont aujourd'hui une réalité quotidienne. En oncologie notamment, l'examen radiologique permet à la fois de suspecter un diagnostic et de suivre l'efficacité des traitements. Depuis quelques années, l'application de techniques mathématiques de traitement des images laissent apparaitre la possibilité de caractériser de manière plus fine les images anormales et d'en extraire des valeurs quantitatives possiblement prédictives, diagnostiques ou encore pronostiques de la maladie.

Dans ce chapitre, nous parcourrons brièvement la problématique clinique et les besoins radiologiques associés. Le concept de Radiomique sera introduit avant de définir les enjeux de la thèse de sciences.

1/ Évaluation des traitements anti cancéreux par imagerie

1.1 L'évaluation en routine clinique

Les patients suivis en oncologie qui reçoivent des traitements systémiques sont surveillés par des bilans réguliers afin d'évaluer leur tolérance et leur efficacité. On considère en effet que le traitement doit être poursuivi tant que sa toxicité est supportable et que les lésions tumorales ne parviennent pas à progresser en taille malgré la pression du traitement.

L'identification d'une progression repose sur un faisceau d'arguments : l'altération de l'état général du patient, l'augmentation de biomarqueurs sanguins lorsqu'ils existent (par exemple l'élévation du Prostate Specific Antigen dans les cancers de la prostate), mais surtout sur un argument objectif déterminé par l'imagerie qui est l'augmentation en taille des lésions tumorales Ce constat est la base des critères d'évaluation tumorale dont l'utilisation est recommandée en routine dans le cadre des tumeurs solides : les critères RECIST (Response Evaluation Criteria In Solid Tumours) version 1.1[1].

En pratique, le radiologue sélectionne entre 1 et 5 lésions « cibles » sur l'examen d'imagerie utilisé pour le suivi, et qui est dans l'immense majorité des cas un examen tomodensitométrique, du fait de sa disponibilité. Ces lésions sont mesurées juste avant le début d'un traitement systémique donné (cet examen est appelé « scanner baseline »). La somme des mesures des lésions constitue la référence pour le suivi. Sur chaque examen suivant, le radiologue mesure les mêmes lésions et évalue par un pourcentage l'évolution de la maladie. La progression est définie par une augmentation de plus de 20% de la somme par rapport au moment de la meilleure réponse (ou Nadir), et/ou l'apparition d'une ou plusieurs nouvelles lésions tumorales.

Ces critères ont été largement adoptés par la communauté internationale, car ils répondaient à un vrai besoin de standardisation de l'évaluation tumorale et qu'ils étaient suffisamment simples pour être pratiqués en routine clinique. Cette approche quantitative est également utilisée dans le cadre de l'évaluation des essais thérapeutiques décrit ci-dessous.

1.2 L'évaluation dans le cadre des essais thérapeutiques

L'accélération de la recherche dans le domaine des traitements anticancéreux a permis le développement d'une multitude d'approches nouvelles dont il fallait valider l'efficacité et identifier la place dans la stratégie thérapeutique.

Les essais thérapeutiques préalables à la validation d'une nouvelle drogue sont standardisés :

une nouvelle molécule doit être testées par des études successives. Le principal argument d'efficacité d'un nouveau traitement en cancérologie repose sur la « survie globale » c'est-à-dire le temps passé entre le début du traitement et le décès du patient. Si la survie globale est meilleure au plan statistique dans un groupe de patients qui reçoit le traitement expérimental, comparativement à un autre groupe de patients traité par le traitement habituel, le traitement expérimental est validé.

Cette stratégie n'est pas toujours possible, notamment chez des patients qui ont des cancers d'évolution lente : il est en effet difficile « d'attendre » le décès des patients pour conclure à l'efficacité d'un traitement.Afin d'accélérer le processus de développement de nouvelles drogues, on est souvent amené à utiliser un substitut (« surrogate ») à la survie globale, qui est la « survie sans progression ».

Les critères RECIST ont initialement été développés pour standardiser l'évaluation des traitements dans le cadre précis des essais thérapeutiques. L'évaluation de l'efficacité est réalisée par des scanners successifs et on définit la « survie sans progression » par le temps passé entre le début du traitement et la progression radiologique selon les critères RECIST : si la survie est statistiquement significativement supérieure dans le groupe expérimental, on considère que le nouveau traitement est a priori meilleur et il peut être développé à l'échelle clinique. Toutefois les critères RECIST semblent insuffisant dans le contexte de la médecine personnalisée lorsqu'il s'agit d'évaluer les nouvelles thérapies ciblées.

1.3 Thérapies ciblées et Médecine personnalisée

Les progrès dans l'identification et la compréhension du fonctionnement des cellules tumorales à l'échelle génomique et métabolique, ont mené au développement de traitements de plus en plus spécifiques appelés thérapies ciblées. En effet, si l'on parvient à identifier au sein du tissu tumoral une anomalie particulière, il devient possible de cibler cette anomalie par un traitement spécifique afin de détruire les cellules tumorales préférentiellement. Ce principe est la base de la médecine personnalisée (figure 1).

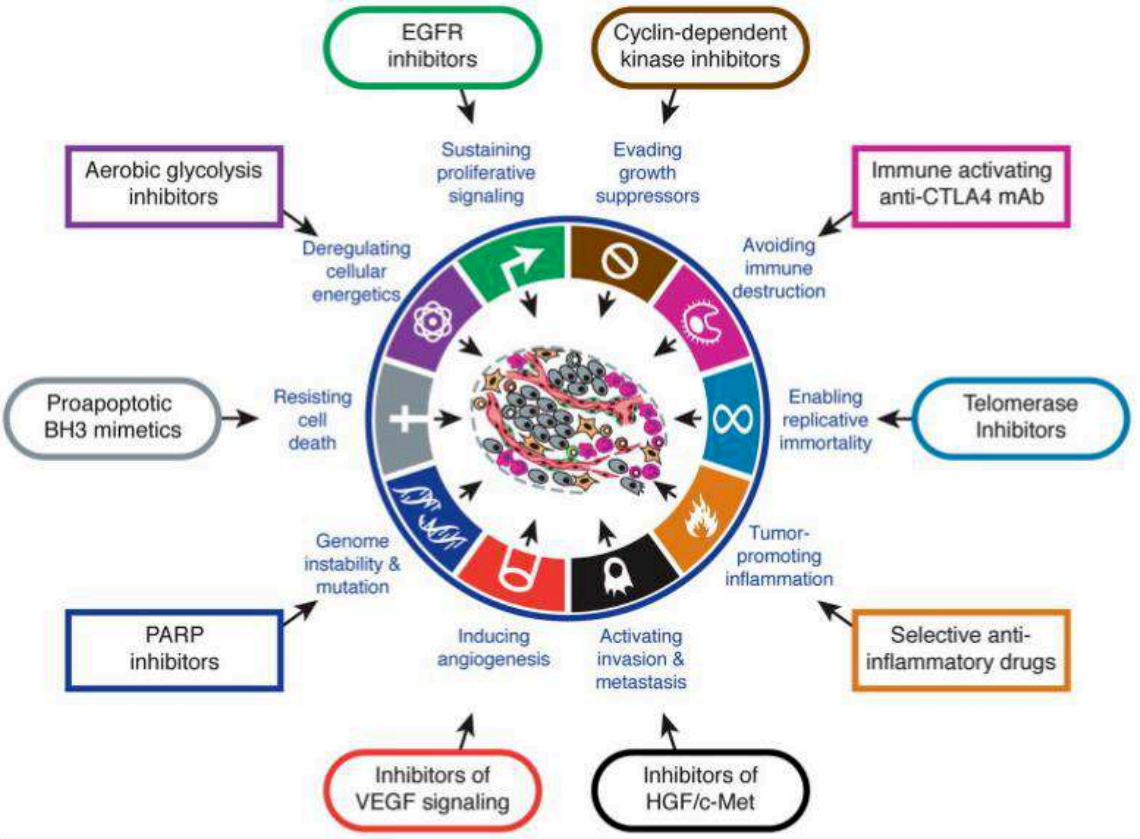


Figure 1 : les thérapies ciblées : illustration de différentes voies de développement selon l'anomalie trouvée au sein des cellules cancéreuses (from Hanahan and Weinberg, Cell 2011[2])

Quelles sont les conséquences du développement de ces nouveaux types de traitements sur l'évaluation de la réponse tumorale ?

Le développement de ces nouvelles thérapies ciblées ayant des mécanismes d'action différents, a provoqué une remise en cause de ces critères d'évaluation classiques. Ainsi par exemple les antiangiogéniques qui ciblent la néovascularisation tumorale n'entrainent parfois pas immédiatement une diminution de la taille des lésions tumorales mais une dévascularisation de celles-ci[3]. Dernièrement, l'apparition d'une nouvelle stratégie thérapeutique qui repose sur le blocage de l'inhibition des lymphocytes en charge de détruire les cellules cancéreuses par les inhibiteurs de « check points » moléculaires, a également montré les limites d'une évaluation purement basée sur l'évolution de la taille des lésions tumorales[4].

La possibilité de disposer de plusieurs types de traitement pour lutter contre la maladie cancéreuse est un progrès indéniable, mais de nouvelles questions autour des indications précises, de l'efficacité réelle et du coût des traitements se posent de plus en plus.

On comprend donc l'intérêt croissant de trouver de nouveaux biomarqueurs radiologiques de l'efficacité des traitements qui ne reposent pas uniquement sur la notion de taille des lésions au scanner.

2/ Définition des biomarqueurs et leur évaluation

Au sens large, un biomarqueur peut être défini par toute caractéristique moléculaire, histologique, radiographique ou physiologique pertinente en termes d'utilité clinique, c'est-à-dire que sa mesure est un indicateur de processus biologiques normaux ou pathologiques ou de la réponse à une exposition ou une intervention thérapeutique[5]. Des biomarqueurs de tous types sont aujourd'hui largement utilisés en cancérologie, que ce soit pour stadifier et caractériser le cancer, cibler les traitements chirurgicaux et radiothérapeutiques ou prédire et surveiller l'efficacité thérapeutique.

La recherche actuelle de nouveaux biomarqueurs de l'efficacité thérapeutique se focalise d'une part sur la recherche de marqueurs biologiques moins invasifs à partir de biopsies liquides, de cellules tumorales circulantes et d'ADN tumoral circulant[6], et d'autre part sur la recherche de biomarqueurs d'imagerie innovants.

On a vu ainsi des études sur l'analyse de paramètres en IRM (coefficient apparent de diffusion notamment) ou en médecine nucléaire avec différents traceurs, pour diverses indications (divers types tumoraux, divers traitements[7,8]. La limite récurrente de ces études était principalement l'absence de reproductibilité de la mesure, dépendant très largement des conditions d'acquisition des images. De plus ces techniques n'étant pas facilement disponibles, leur utilisation hors études était trop complexe à mettre en œuvre.

La conjonction des développements en instrumentation et analyse numérique des images médicales et des enjeux cliniques importants afin d'identifier de nouveaux biomarqueurs d'imagerie est à l'origine d'un nouveau domaine exploratoire en radiologie, la radiomique, dont les principes vont être rappelés dans la partie suivante.

3/ La radiomique

Parallèlement aux avancées dans l'identification des nouvelles cibles thérapeutiques, une autre révolution a eu lieu dans le domaine de l'imagerie, avec la généralisation de la numérisation des examens médicaux ouvrant le champ à de nouvelles perspectives. Une nouvelle discipline a ainsi vu le jour : la radiomique, qui se base sur les images médicales qui ne sont plus considérées comme des images simples, mais aussi comme des données quantifiables que l'on peut traiter avec des approches mathématiques plus complexes[9](figure 2).

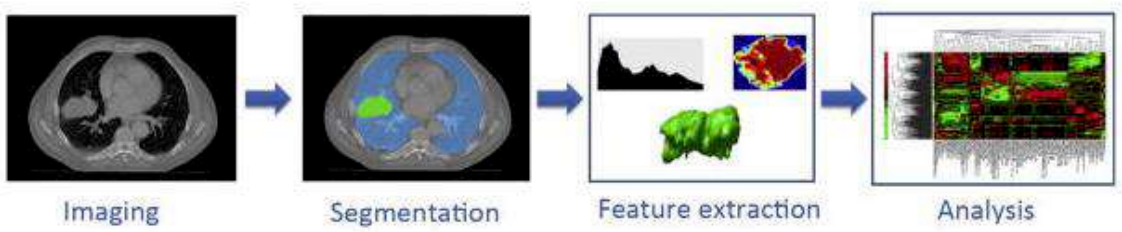


Figure 2 : Représentation schématique du processus de radiomique : partant d'une image médicale (ici un scanner en coupe axiale), la lésion d'intérêt (en vert) est segmentée puis les valeurs en unité Hounsfield des pixels/voxels constitutifs de l'image sont analysés et on extrait des indices de texture. On utilise ces valeurs quantitatives pour des analyses statistiques (image extraite de l'article[10]).

La radiomique s'est développée à partir de l'analyse d'image d'une région d'intérêt contenue dans l'examen numérisé. Cette région d'intérêt est analysée en considérant i) la répartition de l'intensité des pixels/voxels et on parle alors des indices en rapport avec l'histogramme des intensités ; ii) l'organisation spatiale relative de chacun des pixels (/voxels) via le calcul de matrice mathématique de la valeur numériques des pixels/voxels et on parle d'analyse de texture. Les principaux indices sont définis plus précisément dans l'annexe 1.

On peut donc décrire l'hétérogénéité d'une image (par exemple, l'image d'une tumeur) par la quantification des pixels de différentes intensités, par leur organisation dans l'espace, par une analyse chiffrée de la forme, des contours, du volume tumoral. On accède donc à un très grand nombre d'indices mathématiques décrivant l'image. L'analyse de ces données chiffrées est ensuite agrégée et traitée via des analyses statistiques et parfois via les statistiques complexes des techniques d'intelligence artificielle (machine learning).

Aujourd'hui, l'objectif est d'identifier un indice ou un ensemble d'indices d'hétérogénéité comme étant un biomarqueur de l'efficacité des traitements personnalisés, ou bien prédictif de la réponse ou encore de la toxicité. L'enjeu est finalement d'accéder à une sorte de biopsie virtuelle où l'image radiologique permettrait d'accéder au phénotype de la tumeur pour anticiper la stratégie thérapeutique. Dans ce contexte, il existe de nombreuses études cliniques visant à identifier des biomarqueurs de l'hétérogénéité tissulaire à partir de traitements d'images plus ou moins novateurs incluant les analyses de forme et de texture. Toutefois, les publications scientifiques apparues sur ces sujets mettent en lumière l'absence de consensus sur les définitions de la texture en imagerie médicale, et sur l'absence de standardisation des analyses, qui dépend principalement du matériel dont on dispose[11]. En effet, l'image d'une tumeur, obtenue après un examen tomodensitométrique, est le résultat combiné d'une chaine de détection de rayons X avec des phénomènes d'interactions

physiques complexes, associée à une chaine de post-traitement de données constructeur-dépendant et enfin de l'image d'un patient particulier. Appréhender ces variations est aujourd'hui nécessaire pour exploiter les indices d'hétérogénéité à l'échelle des études multicentriques, condition nécessaire pour valider des biomarqueurs d'imagerie.

Dans ce contexte, mon travail de thèse s'inscrit initialement dans le domaine de l'imagerie TDM quantitative. En particulier, il vise à comprendre en amont **les sources de variation des indices d'hétérogénéité inter et intra-scanner**, pour à terme, **proposer une méthode de calibration robuste et standardisée en TDM**. L'enjeu clinique est de pouvoir proposer au radiologue une méthode de caractérisation fine de la tumeur par imagerie pour renforcer l'évaluation précoce des thérapies et permettre ainsi d'orienter une stratégie thérapeutique personnalisée.

En parallèle des avancées dans le domaine de la radiomique, de nouveaux traitements anticancéreux, faisant appel non plus à la destruction directe des cellules tumorales (soit par des chimiothérapies cytotoxiques ou cytostatiques, soit par des thérapies ciblées) mais à la modification des interactions entre cellules tumorales et cellules immunitaires (immunothérapie) ont été découverts. Il s'agit d'une révolution dans le traitement des patients atteints de maladies métastatiques, en particulier dans le mélanome et le cancer bronchique. En effet, rapidement après les premiers essais thérapeutiques réalisés dans ces deux maladies, des autorisations de mises sur le marché ont été obtenues et ces traitements font désormais partie des premières armes disponibles pour les patients[12,13]. L'évaluation de la réponse à ce nouveau type de traitement a été rapidement compliquée par la constatation de réponses atypiques : pseudoprogression (augmentation initiale des lésions suivies d'une stabilisation ou d'une réponse objective) et hyperprogression (accélération de la cinétique tumorale au début du traitement)[14,15]. Dans le cadre de ma thèse, je me suis donc également intéressée à ces nouveaux challenges de l'évaluation tumorale.

Chapitre 2: Répétabilité et reproductibilité des indices de texture en imagerie tomodensitométrique.

En oncologie, un essai thérapeutique peut être ouvert dans plusieurs centres cancérologiques différents (essais multicentriques). L'évaluation de la réponse en imagerie est standardisée grâce aux critères RECIST dont nous avons parlé au chapitre 1. De ce fait, on, demande aux patients de réaliser des examens scanographiques réguliers, qu'ils peuvent faire dans n'importe quel site qui dispose d'une machine ; les images numérisées sont ensuite envoyées dans le centre de l'investigateur principal. Cela simplifie le parcours du patient et permet le recrutement de patients dans une aire géographique plus large. La grande variété des sites d'imagerie pour le suivi des patients en cancérologie a très vite soulevé la question de la possibilité de l'approche multicentrique de l'analyse de la texture des images tomodensitométriques ; en particulier, l'enjeu était d'évaluer la robustesse des indices de texture en fonction des paramètres d'acquisition des différents scanners. Cette question est légitime, dans la mesure où la texture est directement liée à la nature de l'image, et donc à la nature physique du signal. O'Connor (2017) définit la feuille de route pour établir la robustesse d'un biomarqueur en imagerie avec l'étude de la variabilité quantitative du biomarqueur lors d'expériences de répétabilité et reproductibilité[16]

Nous avons ainsi développé et mis en œuvre une méthodologie, décrite dans ce chapitre, intégrant la conception d'un fantôme original mimant la texture homogène et hétérogène des tissus en unité Hounsfield. Après avoir décrit le matériel développé et utilisé, nous présenterons la variabilité des indices dans un contexte de la **répétabilité** des mesures, c'est à dire des mesures faites successivement dans les mêmes conditions. L'étude de la **reproductibilité** de la mesure des indices dans différentes conditions expérimentales sera également développée.

1/ Les points clés des études méthodologiques

Avant de présenter la méthodologie et les résultats des deux études de répétabilité et reproductibilité, je vais détailler les choix faits sur deux aspects essentiels et communs aux deux études : i) un objet test appelé fantôme devant mimer un tissu hétérogène, avec des valeurs d'intensité en unité Hounsfield réalistes et confectionné dans un matériau durable dans le temps pour assurer un grand nombre de tests ; ii) un logiciel d'extraction et de détermination des indices de texture qui doit être d'une très grande fiabilité.

1.1 L'objet test en imagerie : le fantôme

Le cahier des charges pour la conception d'un fantôme utilisable pour l'étude de la variabilité de la répétabilité et de la reproductibilité des indices de texture en tomodensitométrie était de :

- Permettre l'acquisition de plusieurs zones de texture différentes, homogènes ou hétérogènes, avec des valeurs d'intensité en Unité Hounsfield (UH) réalistes et stables dans le temps pour être réutilisables
- Permettre de maitriser la disposition spatiale des inserts de textures différentes
- Permettre de transposer à l'identique les manipulations sur différents sites/ différents appareils / par différentes équipes

A ce jour, il n'existe pas de tel fantôme validé et commercialisé. Nous avons donc développé un fantôme visant à répondre à l'ensemble des critères sus-cités en s'appuyant sur un équipement déjà commercialisé par CIRS pour l'analyse de la qualité image : il s'agit du « 062M Electron Density Phantom ». Comme illustré dans la figure 3, celui-ci est constitué d'une couronne comportant des alvéoles, dans lesquelles on peut disposer, au choix, soit des inserts homogènes de densités connues fournis par le constructeur, soit des inserts « faits-maison » grâce à des seringues dont les dimensions s'adaptent aux alvéoles. Nous avons donc conçu deux inserts hétérogènes contenant un mélange aléatoire d'Ecoflex® (BASF, Ludwigshafen, Germany), qui est de la famille des silicones, avec des fragments de gel d'agar-agar carboné, ces matériaux étant choisies pour leur imputrescibilité impliquant une bonne tenue mécanique dans le temps. Deux inserts hétérogènes ont été conçus selon deux mixtures différentes de densités hétérogènes le plus proche possible d'une densité de tumeur « humaine » : ils diffèrent l'un de l'autre par la répartition interne des zones carbonées. Afin de s'approcher des conditions expérimentales de certaines études décrites dans la littérature, nous avons aussi initialement conçu des inserts en bois et en liège[17], et un insert contenant de l'Ecoflex® mélangé de manière aléatoire avec cette fois-ci des billes de polystyrène[18].

Figure 3 : Photographies du fantôme utilisé dans les expériences des études de répétabilité et reproductibilité

Dans la suite de l'étude, nous avons considéré entre 3 et 6 inserts numérotés de la façon suivante :

C1 : homogène simulant la densité musculaire avec un pic d'intensité mesuré à 48 UH. Cet insert sera considéré comme l'insert homogène de référence dans tout ce travail de thèse.

C2 : Matériaux hétérogène avec un mélange de silicone et d'éléments carbonés mimant les tissus tumoraux.

C3 : très similaire à C2 avec une concentration plus élevée d'éléments carbonés.

C4 : hétérogène avec un mélange de silicone et de polystyrène.

C5 : hétérogène en liège permettant de proposer une image plus « texturée » en CT.

C6 : hétérogène en bois sipo permettant de proposer une image plus « texturée » en CT.

Les valeurs en unité Hounsfield (UH) de chacun de ces six inserts ont été obtenues à partir des images scanographiques acquises avec un Scanner CT General Electric Discovery CT750 HD (cf. figure 4).

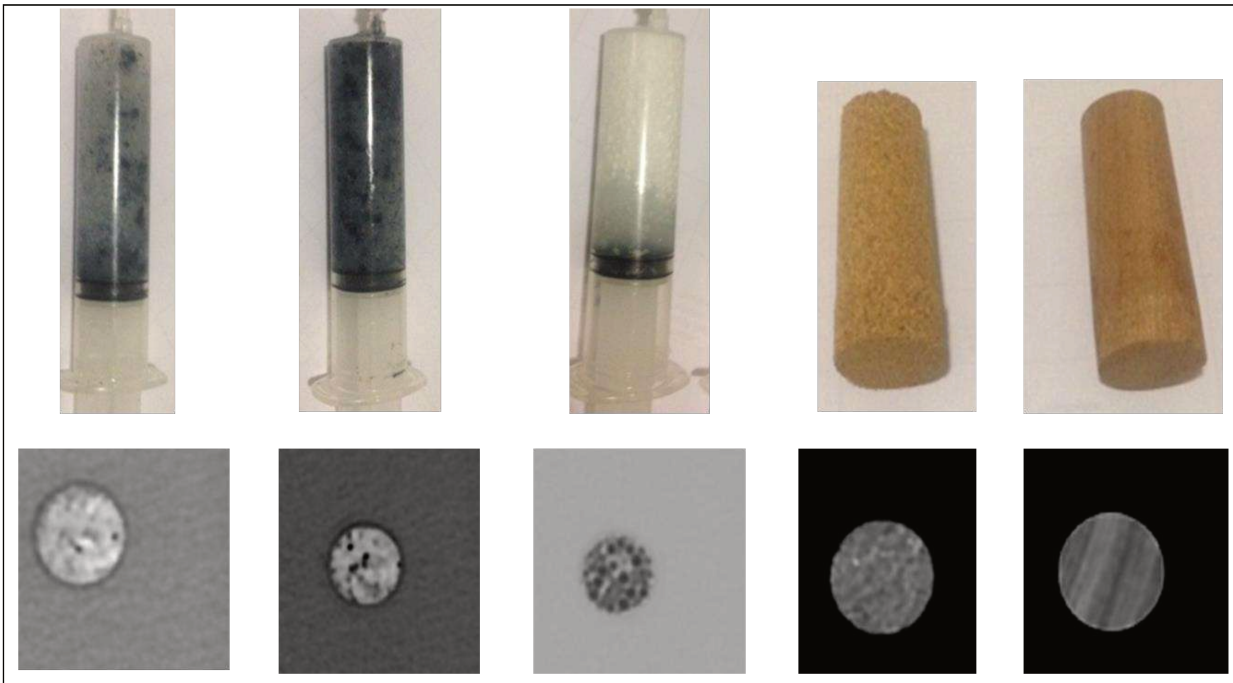


Figure 4 : photographies des inserts C2 à C6 et leur image scannographique dans le plan axial.

Il apparaît que la distribution des intensités de l'insert C1 de référence est bien autour de la valeur 48 UH comme indiqué par le constructeur. Concernant les valeurs centrales des distributions des inserts C4, C5 et C6, celles-ci sont particulièrement éloignées des ordres de grandeur des tissus, quelle que soit leur nature. Plus précisément, les inserts C5 et C6 composés de liège et de bois présentent des densités très négatives (respectivement densité moyenne -390 UH (-470 à -330) et -776 UH (-888 à – 697 UH) , éloignées de ce que l'on peut rencontrer en condition médicale, de même que l'aspect visuel des images, avec des lignes régulières pour le bois et des « points réguliers » pour le liège. Enfin l'insert composé d'Ecoflex mélangé à des billes de polystyrène comprenait des densités très différentes et surtout des bulles d'air persistantes qui rendaient l'analyse impossible. Concernant la répartition des valeurs en unités Hounsfield des pixels dans les régions d'intérêt issues des images des inserts C2 et C3 hétérogènes (Ecoflex + agar-agar carboné), celle-ci forme une gaussienne de valeurs moyennes se rapprochant des densités trouvées dans le corps humain et sont légèrement différents l'un de l'autre.

Dans la suite de ce chapitre, nous focaliserons nos études sur les inserts C1 homogène et sur les inserts C2, C3 hétérogènes composée d'Ecoflex et agar-agar carboné. L'histogramme de répartition des valeurs des pixels des 3 inserts est illustré sur la figure 5.

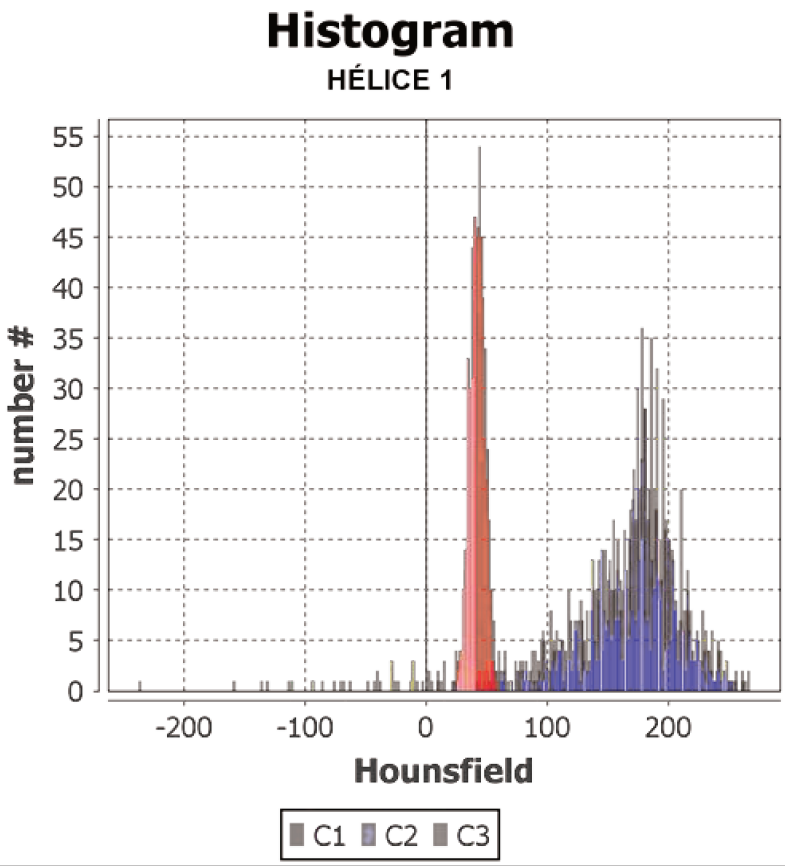


Figure 5 : Histogramme des 3 régions d'intérêt C1, C2 et C3

1.2 Logiciel d'analyse

Le logiciel d'extraction et d'analyse des indices peut être une composante de la variabilité quantitative des indices de texture avec des définitions mathématiques parfois différentes. Dans le cadre de notre étude, nous avons choisi d'utiliser le logiciel LIFEx, disponible en libre accès, fiable, et qui intègre les dernières recommandations internationales sur la détermination des indices de texture (Image Biomarker Standardization Initiative guidelines)[19](figure 6).

La région d'intérêt définie pour extraire les indices de texture est une zone d'intérêt bidimensionnelle de diamètre fixe : elle a été définie sur les images tomodensitométriques obtenues sur le fantôme, pour chaque insert.

Par ailleurs le logiciel LIFEX offre la possibilité de modifier le pas de discrétisation. Celui-ci correspond au ré-échantillonnage des intensités des pixels présents dans la ROI par niveau de gris, c'est-à-dire que pour une discrétisation de 1, il y a autant de niveaux de gris qu'il y a d'intensités, mais pour une discrétisation de 2 il y en a deux fois moins etc… Dans le cadre de notre étude, le pas de discrétisation était fixé de -1000 UH jusqu'à 1000 UH par pas de 10 UH (soit 200 niveaux de gris).

Enfin, le logiciel fournit les résultats de l'analyse de texture de 34 indices définis dans l'annexe 1 et que l'on peut décomposer en deux familles distinctes :

1) L'analyse statistique du premier ordre qui regroupe les approches statistiques descriptives comme la moyenne, la médiane, l'écart-type, mais également des indices descriptifs de l'histogramme des intensités appelés Skewness, Kurtosis, EntropyH, EnergyH (la lettre H ici est ajoutée pour différencier des autres indices de même nom mais extraits à partir de matrices) ;
2) L'analyse statistique d'ordre deux mettent en jeu plusieurs pixels simultanément et permettent donc de conserver les relations spatiales entre pixels. Avec LIFEX, nous avons la possibilité d'extraire des indices de cette façon à partir de quatre matrices mathématiques :**la matrice de co-occurrence** (GLCM, 6 indices : Homogeneity, Energy, Contrast, Correlation, Entropy Dissimilarity), de **la matrice dite Grey Level Run Length** (GLRLM avec 11 indices : SRE/LRE, LGRE / HGRE, SRLGE/ SRHGE, LRLGE/LRHGE, GLNU/RLNU, RP), **la matrice dite Neighborhood Grey Level difference** (NGLDM, avec 3 indices Coarseness, Contrast, Busyness) et **enfin la matrice Grey-Level Size Zone** (GLSZM, 11 indices : SZE / LZE, LGZE / HGZE, SZLGE / SZHGE, LZLGE/ LZHGE, GLNUz / ZLNU, ZP).

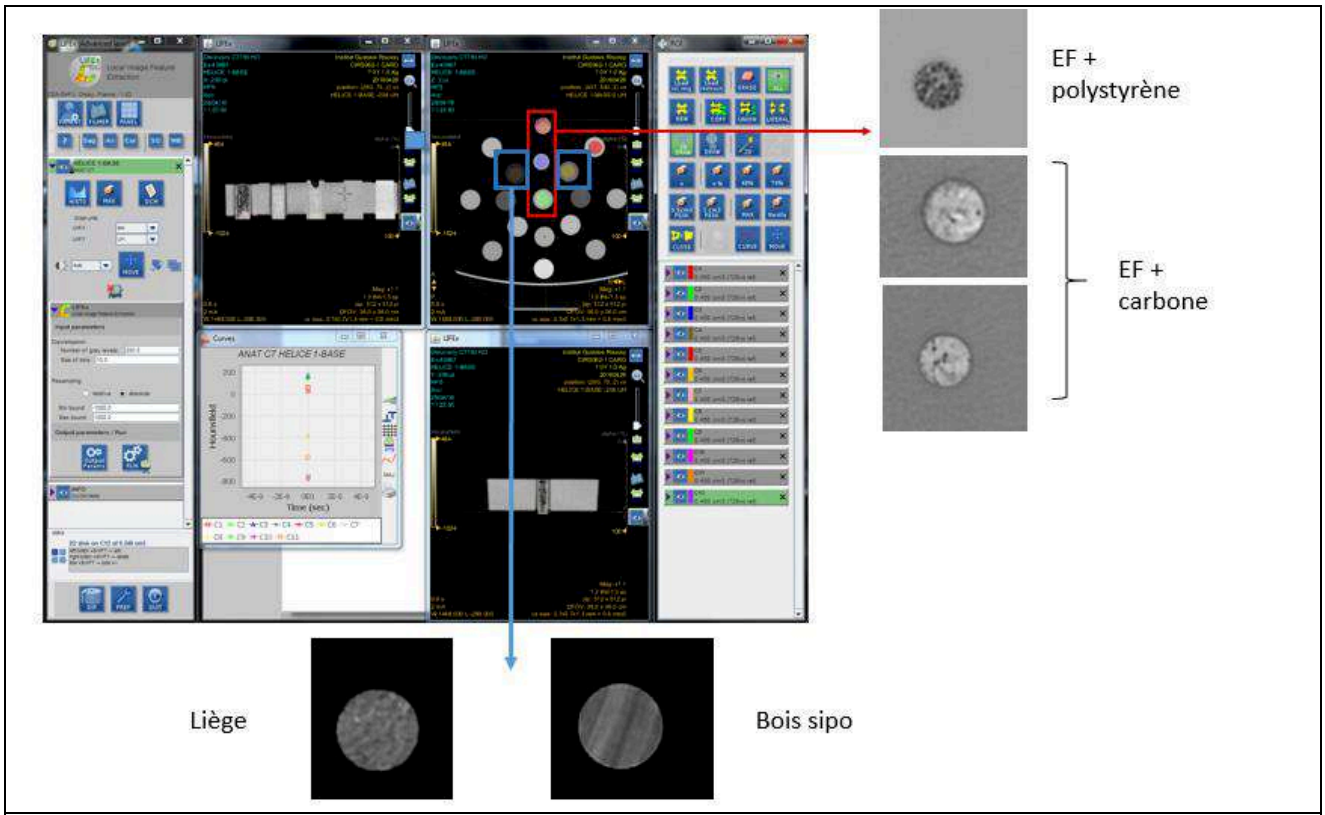


Figure 6 : Capture d'écran de la visualisation rendue par le Logiciel LIFEx, et agrandi de l'aspect visuel des différentes zones d'intérêt au scanner.

2/ Étude de la répétabilité de la mesure des indices de texture

L'objectif de cette étude est de déterminer de manière quantitative la variabilité des indices de texture lors d'acquisition d'images TDM répétitives et ce, sans a priori sur leur prédisposition à être des biomarqueurs. La capacité des indices à dissocier les milieux hétérogènes versus homogène sera également explorée. Cette étude vise ainsi à identifier les indices les plus robustes à la répétabilité.

2.1 Méthode

2.1.1 Acquisitions

Nous avons dans un premier temps réalisé 8 acquisitions identiques sur le fantôme à partir de la même machine (General Electric Discovery CT750 HD , 64 détecteurs). Nous avons choisi un protocole standard utilisé pour l'imagerie du thorax, en bloquant le système d'auto-milliampérage, afin d'avoir exactement les mêmes conditions expérimentales sur toutes les coupes.

Les paramètres précis d'acquisition étaient les suivants :

- Tension du tube : 120 kVp
- Intensité du courant du tube: 150 mAs
- Epaisseur de coupe : 1.25 mm
- Pitch : 1.375
- Temps de rotation : 0.6 s
- Champs de vue : 360 mm
- Filtre : standard
- ASIR (index de reconstruction propre à General Electric) : 0%.

Dans un second temps, nous avons modifié uns à uns les paramètre d'acquisition afin d'obtenir 39 nouvelles séries d'images du fantôme (tableau 1). Cette étape nous permettra d'identifier de potentielle redondance entre les indices.

Acquisition	kVp	mAs	Epaisseur de coupe (mm)	pitch	Temps de rotation (s)	Champ de vue (mm)	ASIR (%)	Filtre
base	120	150	1.25	1.375	0.6	360	0	standard
2	80	150	1.25	1.375	0.6	360	0	standard
3	100	150	1.25	1.375	0.6	360	0	standard
4	140	150	1.25	1.375	0.6	360	0	standard
5	120	100	1.25	1.375	0.6	360	0	standard
6	120	200	1.25	1.375	0.6	360	0	standard
7	120	250	1.25	1.375	0.6	360	0	standard
8	120	300	1.25	1.375	0.6	360	0	standard
9	120	350	1.25	1.375	0.6	360	0	standard
10	120	400	1.25	1.375	0.6	360	0	standard
11	120	500	1.25	1.375	0.6	360	0	standard
12	120	150	0.625	1.375	0.6	360	0	standard
13	120	150	2.5	1.375	0.6	360	0	standard
14	120	150	3.75	1.375	0.6	360	0	standard
15	120	150	5	1.375	0.6	360	0	standard
16	120	150	1.25	0.516	0.6	360	0	standard
17	120	150	1.25	0.984	0.6	360	0	standard
18	120	150	1.25	1.375	0.4	360	0	standard
19	120	150	1.25	1.375	0.5	360	0	standard
20	120	150	1.25	1.375	0.7	360	0	standard
21	120	150	1.25	1.375	0.8	360	0	standard
22	120	150	1.25	1.375	0.9	360	0	standard
23	120	150	1.25	1.375	1	360	0	standard
24	120	150	1.25	1.375	0.6	160	0	standard
25	120	150	1.25	1.375	0.6	260	0	standard
26	120	150	1.25	1.375	0.6	460	0	standard
27	120	150	1.25	1.375	0.6	360	50	standard
28	120	150	1.25	1.375	0.6	360	10	standard
29	120	150	1.25	1.375	0.6	360	20	standard
30	120	150	1.25	1.375	0.6	360	30	standard
31	120	150	1.25	1.375	0.6	360	40	standard
32	120	150	1.25	1.375	0.6	360	60	standard
33	120	150	1.25	1.375	0.6	360	70	standard
34	120	150	1.25	1.375	0.6	360	80	standard
35	120	150	1.25	1.375	0.6	360	90	standard
36	120	150	1.25	1.375	0.6	360	100	standard
37	120	150	1.25	1.375	0.6	360	0	soft
38	120	150	1.25	1.375	0.6	360	0	detail
39	120	150	1.25	1.375	0.6	360	0	bone
40	120	150	1.25	1.375	0.6	360	0	lung

Tableau 1: Résumé des paramètres d'acquisitions scannographiques réalisées sur le fantôme (en jaune les éléments variables)

2.1.2 Analyse statistique

Les 8 acquisitions identiques ont permis d'extraire 8 tableaux de données décrivant la valeur des 34 indices de texture, pour chacune des 3 régions d'intérêt. A partir de ces données, nous allons déterminer trois critères : la répétabilité des indices, leur capacité discriminatoire entre les milieux homogènes et hétérogènes et enfin leur corrélation éventuelle.

La répétabilité de la mesure a été estimée par le coefficient de variation (CV) selon la formule suivante :

$$cv = \frac{\sigma}{\mu}$$

où σ correspond à la déviation standard et μ à la valeur moyenne de l'échantillon.

Pour identifier les indices dits robustes à la répétabilité, nous avons décidé dans cette étude qu'un CV inférieur à 20% était acceptable dans une routine clinique. Un seuil de variation des coefficients de variation concernant les biomarqueurs médicaux « acceptable » n'existant dans aucune recommandation, il a été fixé de manière arbitraire[16].

La capacité des indices de texture à discriminer milieux homogènes et hétérogènes est un élément essentiel dans un contexte de suivi de l'efficacité thérapeutique. Aussi, nous nous sommes intéressés à comparer les valeurs des indices selon les trois inserts en vue d'identifier les indices permettant de différencier C1 (homogène) d'une part et C2/C3 (hétérogènes) d'autre part. Il était également important qu'un indice permette de différencier C2 et C3, les deux inserts hétérogènes visuellement différents mais de densités proches. Pour quantifier au mieux cette capacité discriminatoire, nous avons considéré le test non-paramétrique de Mann-Whitney.

Enfin, pour évaluer une éventuelle redondance entre les indices, nous avons utilisé les 40 séries acquises avec différents paramètres d'acquisition et appliqué une corrélation de Pearson. Une matrice de corrélation a servi à explorer les relations 2 à 2 entre les paramètres retenus et une analyse en composante principale a été utilisée pour représenter de manière visuelle simple les indices retenus et leur modèle de corrélation.

2.2 Résultats

2.2.1 Répétabilité de la mesure des 34 indices de texture

L'ensemble des coefficients de variations (CV) calculés pour chaque indice de texture issu des trois inserts C1, C2 et C3 sont résumés dans le tableau 2. Il en ressort que pour 32 indices, les coefficients de variation varient entre 0.14% et 26.22% avec une médiane de 2.25%. En revanche il apparaît que 2 indices : skewness (histogramme) et buzyness (NGLDM), ont des CV très élevés. Ces fortes variations s'expliquent par leur définition mathématique qui implique des valeurs moyennes, dénominateur de CV, proches de zéro.

LIFEx	CV	CV	CV
Name of ROI	**C1**	**C2**	**C3**
minValue	15.05	-18.65	-3.72
meanValue	3.6	0.47	0.4
stdValue	3.49	0.9	1.38
maxValue	3.66	1.51	1.27
SkewnessH	-375.13	-4.44	-1.8
KurtosisH	5.11	3.37	2.28
EntropyH	1	0.54	0.27
EnergyH	3.17	3.2	1.45
Homogeneity	1.12	0.88	0.8
Energy	9.63	4.02	2.43
Contrast (co-occurrence)	6.14	2.49	3
Correlation	8.5	0.8	0.38
Entropy	3.5	0.35	0.34
Dissimilarity	3.92	0.84	1.3
SRE	1.85	0.42	0.44
LRE	10.7	2.29	2.25
LGRE	0.3	0.17	0.18
HGRE	0.3	0.17	0.14
SRLGE	2.01	0.38	0.41

LIFEx	CV	CV	CV
Name of ROI	**C1**	**C2**	**C3**
SRHGE	1.73	0.54	0.54
LRLGE	10.82	2.33	2.19
LRHGE	10.58	2.25	2.3
GLNU	2.34	1.22	2.48
RLNU	4.23	1.59	1.54
RP	2.53	0.67	0.59
Coarseness	4.46	2.21	1.43
Contrast (NGLDM)	26.22	3.98	5.99
Busyness	-744.65	-4039.39	-202.31
SZE	8.37	1.68	1.28
LZE	12.21	11.87	17.87
LGZE	0.22	0.28	0.31
HGZE	0.22	0.27	0.25
SZLGE	8.29	1.5	1.35
SZHGE	8.46	1.91	1.44
LZLGE	12.19	11.77	17.51
LZHGE	12.24	11.94	18.14
GLNUz	11.1	3.73	2.65
ZLNU	14.3	5.08	3.59
ZP	6.72	2.56	1.41

Tableau 2 : Résumé des valeurs des coefficients de variations mesurés sur les 3 inserts C1 C2 et C3 pour chaque indice de texture.

2.2.2 Pouvoir de discrimination des indices de texture entre région homogène et hétérogène

Pour identifier les indices de texture permettant de différencier C1, C2 et C3, nous nous sommes appuyés sur les graphiques ci-dessous. On relève trois types de familles d'indices : Ceux qui ne dissocient aucunement les trois inserts ($p>0.05$). C'est le cas de trois indices qui ont des valeurs similaires sur chacune des régions d'intérêt : busyness, GLNU (illustration Figure 7) et SZLGE.

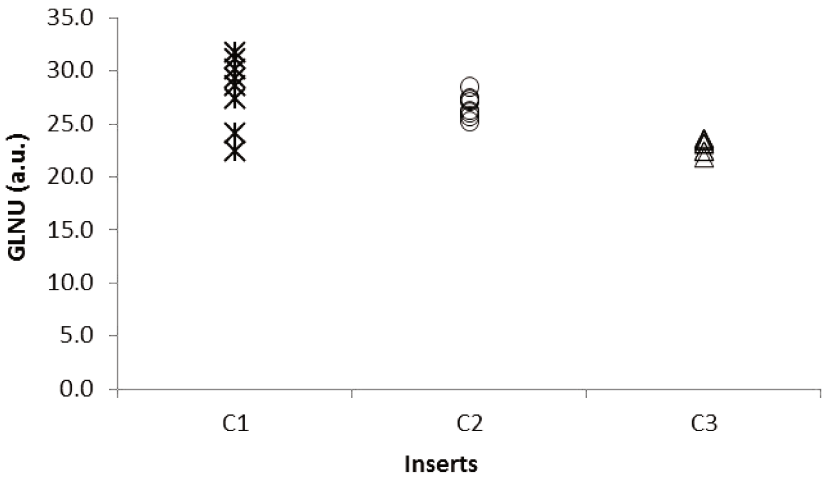


Figure 7: Graphique de la distribution des valeurs de l'indice « GLNU » de la Grey-Level Size Zone Matrix GLSZM pour C1, C2 and C3. Les valeurs de ces indices sont très proches, avec des valeurs communes, ne permettant pas l'exploitation d'un tel indice pour caractériser les différences visuellement évidentes.

Selon la même méthode, nous avons identifié les indices qui ne permettent pas de différencier les 2 régions hétérogènes C2 et C3 ($p>0.05$). C'est le cas de vingt-quatre indices: entropyH et energyH; homogeneity, energy et correlation (illustration sur la figure 8) de la matrice de co-occurrence; SRE, LRE, LGRE, HGRE, SRLGE, SRHGE, LRLGE, LRHGE, RLNU, RP de la Grey-Level Run Length Matrix; et SZE, LZE, LGZE, HGZE, SZHGE, LZLGE, LZHGE, GLNU, ZP de la Grey-Level Size Zone Matrix.

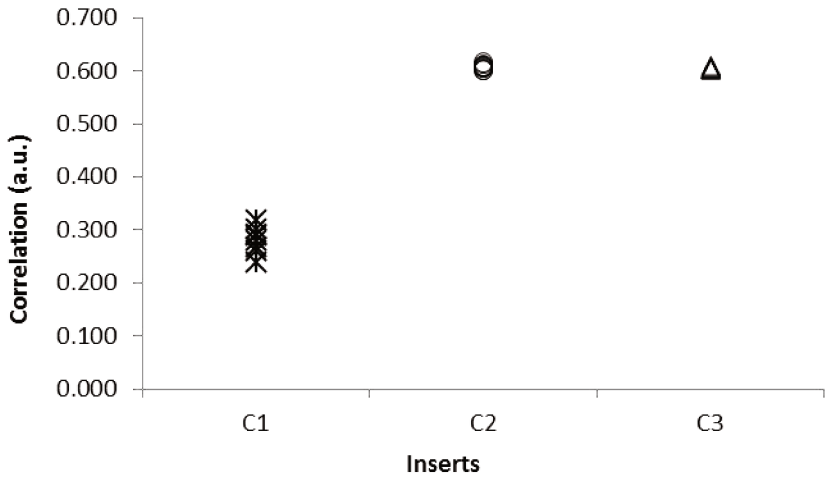


Figure 8: Graphique de la distribution des valeurs de l'indice « Correlation » de la matrice de co-occurrence qui ne montre aucune différence entre C2 et C3.

Au total, les indices capables de distinguer C1 de C2/C3 et C2 de C3 sont au nombre de huit : il

s'agit de skewness (illustration sur la figure 9), kurtosis derives de l'histogramme, contrast et entropy de la matrice de co-occurrence, dissimilarity et contrast de la Neighborhood Grey Level difference matrix, SZE et ZLNU de la Grey-Level Size Zone Matrix. Ces 8 indices sont associés à des p values très significatives (<0.0002), en comparant les valeurs C1 vs C2, C2 vs C3 et C1 vs C3.

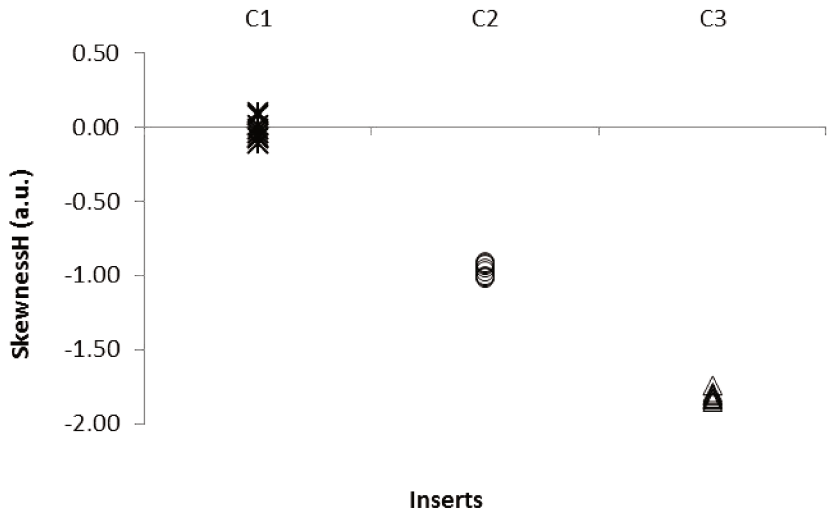


Figure 9: Graphique de la distribution des valeurs de l'indice « Skewness » entre les 3 régions d'interêt démontrant un pouvoir de discrimination entre les 3 zones.

2.2.3. Etude de la corrélation entre les indices

L'étude de la redondance a été réalisée sur les huit indices retenus précédemment et est représentée dans la figure 10 selon l'analyse en composante principale. On constate que skewness et kurtosis semblent indépendants, tandis que les 6 autres indices semblent corrélés entre eux mais que cette corrélation n'est pas mise en évidence pour les 3 zones d'intérêt à la fois. Par exemple, SZE est très corrélé à Contrast pour C2 et C3, mais pas pour C1.

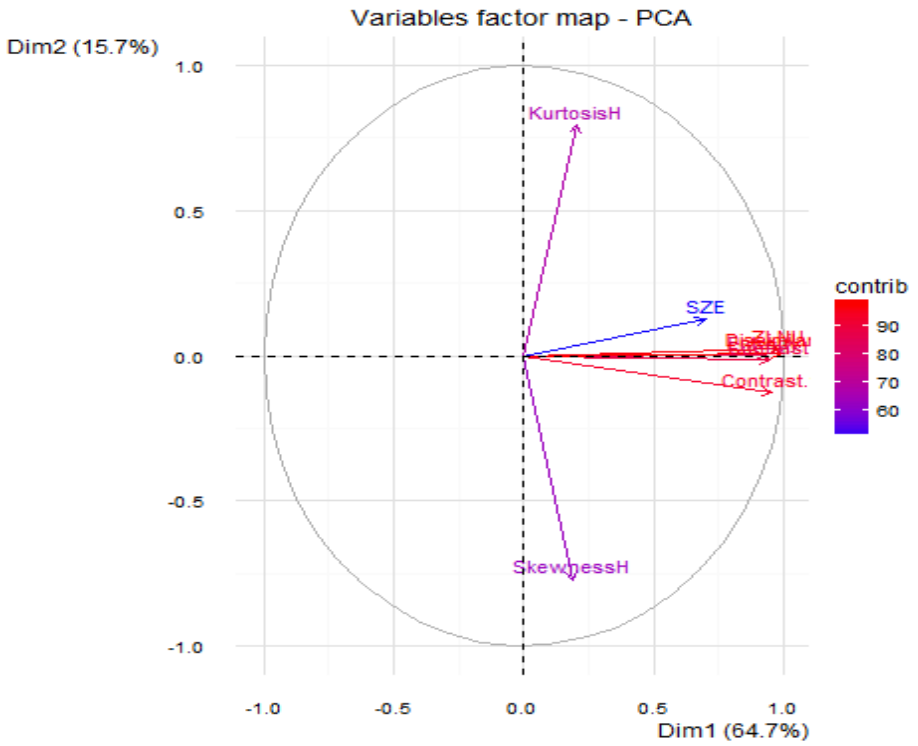


Figure 10: Représentation visuelle des dot plots sous la forme de l'analyse en composante principale montrant la part respective des indices dans l'analyse de redondance.

2.3 Discussion

Il s'agit d'une étude qui visait à explorer la répétabilité des indices de texture dans des conditions expérimentales identiques. Elle a été réalisée dans un contexte où de nombreux articles scientifiques étaient publiés sur l'utilité de la « texture » comme nouveau biomarqueur en imagerie oncologique notamment, alors qu'aucune indication n'était donnée par les auteurs sur les conditions d'acquisition des images utilisées. Il s'agissait alors de la première étude méthodologique sur fantôme en imagerie tomodensitométrique.

Plusieurs auteurs avaient néanmoins soulevé de semblables questions : Fave et al avaient étudié la reproductibilité de 68 indices de texture sur des images de test-retest issues de cone beam CT de 10 patients[20]. Le logiciel utilisé était Ibex. Par ailleurs, ils excluaient 23 indices en raison de leur mauvaise reproductibilité et malheureusement, en dehors de l'indice skewness dérivé de l'histogramme et de contrast dérivé de la matrice de co-occurence, il était impossible de trouver des indices en communs avec notre étude[20]. De plus, les auteurs avaient exclu les indices suivants : energH, correlation de la matrice de co-occurence, LRHGE, SRLGE de la RLM matrice, buzyness et contrast de la NGLDM matrice du fait d'une dépendance de la valeur au volume de la ROI testée. Hunter et al ont étudié la reproductibilité des indices de texture sur des images de scanner sans injection de 56 patients porteurs d'un cancer pulmonaire réalisées 2 fois (test retest qui pouvaient être faits sur 3 machines différentes)[21]. Ils ont exploré 328 indices de texture calculées par le logiciel ibex

(incluant les indices de l'histogramme, de la matrice de co-occurrence et de la Grey Level Run Length Matrix). Ils trouvaient un coefficient de corrélation concordant (supérieur à 0.90) pour 61 à 94.5% des indices. L'analyse de redondance restreignait le nombre d'indices à 23, incluant kurtosis, skewness de l'histogramme, entropy dérivé de la matrice de co-occurence, ce qui est concordant avec notre étude.

Balagurunathan et al étudiaient la reproductibilité de 219 indices appliqués sur des images de test retest de scanner sans injection de 32 patients avec une tumeur pulmonaire et confirmaient la robustesse de l'indice contrast dérivé de la matrice de co-occurence[22].

Dans les conditions expérimentales réalisées, cette étude préliminaire montre :

1/ que certains indices de texture ne sont pas robustes vis-à-vis de la répétabilité évaluée avec le coefficient de variation.

2/ que certains indices de texture sont robustes vis-à-vis de la répétabilité dans les mêmes conditions expérimentales mais ne permettent pas de discriminer une zone d'intérêt homogène d'une zone hétérogène.

3/ que certains indices de texture sont robustes vis-à-vis de la répétabilité mais ne permettent pas de discriminer *dans nos conditions expérimentales* 2 zones d'intérêt hétérogènes proches dans leur composition mais différentes visuellement

4/ que certains indices sont redondants entre eux

Au total, on retient 8 indices sur les 34 étudiés, qui pourraient permettre une analyse quantifiable de zones d'intérêt transposables en pratique clinique.

- Skewness and kurtosis dérivés de l'histogramme
- contrast et entropy de la matrice de co-occurrence
- dissimilarity et contrast de la Neighborhood Grey Level difference matrix
- SZE et ZLNU de la Grey-Level Size Zone Matrix

Les limites de l'étude sont multiples :

- Analyse de zones d'intérêt artificielles qui, malgré des densités proches de celles du corps humain, ne correspondent peut-être pas à des réalités physiopathologiques
- Analyse avec un logiciel unique (LIFEx). L'analyse de texture n'est pas encore un champ scientifique suffisamment mature pour que nous puissions réellement comparer notre méthodologie avec celle d'autres équipes qui s'intéressent au sujet. En particulier, certaines équipes

utilisent d'autres logiciels : soit des matrices mathématiques sur Matlab home-made, soit d'autres logiciel (PyRadiomic, IBEx…). Etant proches des développeurs de LIFEx, nous avons obtenu l'assurance qu'ils sont conformes aux recommandations actuelles (Image Biomarker Standardization Initiative guidelines), en ce qui concerne les formules mathématiques appliquées et la dénomination des indices de texture.

- Analyse de zones d'intérêt bidimensionnelles. Certaines équipes analysent des régions d'intérêt 3D, ce qui ajoute plusieurs autres types de calculs d'indices, notamment des indices relatifs à la forme de la structure étudiée. Nous avons volontairement choisi des régions circulaires, de taille similaire, ce qui ne permet pas d'étudier la variabilité des étapes de segmentation et de contourage manuel/semi-automatique ou automatiques, mais qui permet de limiter les incertitudes dans l'interprétation des résultats.

Publication dans **Medical Physics** : Can we trust the calculation of texture indices on CT images ? A phantom study ; Caramella C, Allorant A, Orlhac F, Bidault F, Asselain B, Ammari S, Jaranowski P, Moussier A, Balleyguier C, Lassau N, Pitre-Champagnat S ; PMID : 29443389 DOI:10.1002/mp.12809)[23].

La préoccupation scientifique autour de la fiabilité de la mesure est également partagée par d'autres groupes de recherche. Ce domaine étant en pleine expansion au moment où nous finalisions l'article scientifique ci-dessus, il nous a semblé nécessaire de reprendre une analyse bibliographique des articles scientifiques traitant de ce sujet. Une recherche pubmed a été faite en utilisant les mots clés "CT or computed tomography" AND "texture or radiomics"). Au moment de cette recherche, 1143 articles répondaient à ces critères. Nous avons retenu les articles qui traitaient de variabilité, reproductibilité ou répétabilité de la mesure des indices de texture, sur fantôme et/ou sur des patients. Nous avons exclu les articles qui étudiaient l'intérêt diagnostique, pronostique ou de prédiction de réponse au traitement, ainsi que ceux qui n'étudiaient que l'histogramme. Au final, au 5 février 2018, 20 articles étaient disponibles.

Le tableau 3 reprend les références bibliographiques, le type d'étude (fantôme ou patient), le nombre de machines analysées, le type de logiciel de texture et les paramètres étudiés.

Reference	Phantom	Patients	Number of CT Devices	Number of Patients	Software	Parameters Studied
Al-Kadi 2009 [7]	No	Lung	2	67	In-house	Repeatability
Balagurunathan 2014 [8]	No	RIDER	2	32	In-house	2D/3D
Berenguer 2018 [9]	Pelvic + CCR copy	No	5	NA	IBEX	Repeatability and redundancy, various acquisition parameters
Buch [10]	In-house	No	1	NA	LIFEx	Tube voltage, current, slice thickness
Caramella 2018 [11]	In-house	No	2	NA	In-house	Repeatability
Fave 2015a [12]	No	NSCLC	? *	20	IBEX	Voltage, current, 2D/3D
Fave 2015b [13]	CCR	NSCLC	19	10	IBEX	Repeatability, CT scanner brand
He 2016 [14]	No	Lung	1	240	In-house	contrast enhancement
Kim 2016 [15]	No	Lung nodule	1	42	In-house	Reconstruction algorithm
Larue 2017 [16]	CCR	NSCLC	9	325	In-house	Repeatability, current, slice thickness
Lu 2016 [17]	No	RIDER	1	32	In-house	Slice thickness, filter
Mackin 2015 [18]	CCR	NSCLC	16	20	IBEX	CT scanner brand
Mackin 2017 [19]	No	NSCLC	1	8	IBEX	Pixel size
Mackin 2018 [20]	CCR	NSCLC	2	107 †	IBEX	Current
Mahmood 2017 [21]	Lung	No	3	NA	IBEX	Filter, CT scanner brand
Midya 2018 [22]	Uniform + anthropomorphic	Abdominal scan	1	1	In-house	Current, reconstruction algorithm
Shafiq-ul-Hassan 2017 [23]	CCR	No	8	NA	In-house	Slice thickness, pixel size
Solomon 2016 [24]	No	Lung, liver, kidney	1	20	In-house	Reconstruction algorithm
Yang 2015 [25]	No	Lung	1	8	IBEX	Contrast enhancement
Zhao 2014 [26]	Thorax	No	1	NA	In-house	Slice thickness, filter

Tableau 3: Caractéristiques des articles sélectionnés. CCR = credence cartridge radiomics phantom; RIDER = Reference Image Database to evaluate Therapy Response; NSCLC = non small cell lung adenocarcinoma. * = les auteurs n'ont pas indiqué le nombre de scanners. †= le nombre de patients n'est pas précisé mais ils proviennent d'une autre étude qui comportait 107 patients[17,20,22–39].

- Dans les articles « fondamentaux » : l'analyse était réalisée sur des fantômes « home-made » non commerciaux, ou des fantômes constitués de matériau dont la texture est très éloignée de la réalité médicale (liège, bois). Ceci était prévisible car il n'existe à l'heure actuelle pas de recommandation sur le type de fantôme à utiliser et chaque équipe a utilisé celui dont elle disposait.
- De plus, divers types de calculs ou matrices de texture étaient appliqués, à nouveau en dehors des recommandations. Ce point a été partiellement résolu par la publication d'un consortium de chercheurs autour de la radiomique en mai 2020[40].
- L'utilisation de machines différentes et de conditions d'acquisition des images différentes, qui n'était pas systématiquement décrites dans les papiers publiés, mais qui était rapportée comme une source importante de variabilité.

Publication dans **Diagnostics (Basel)**. CT Texture analysis challenges: influence of acquisition and reconstruction parameters: a comprehensive review. Espinasse M, Pitre-Champagnat S, Charmettant B, Bidault F, Volk A, Balleyguier C, Lassau N, Caramella C. 2020 Apr 28;10(5):258. doi: 10.3390/diagnostics10050258[41]

3/ Etude de la reproductibilité de la mesure intra et inter machines (modification des conditions expérimentales)

La première partie de ce travail de thèse a déterminé la répétabilité, la discrimination entre des images de densités homogène ou hétérogène, et la redondance de certains indices. Cette première étude avait été réalisée sur la même machine tomodensitométrique dans les exactes mêmes conditions expérimentales. L'étape méthodologique suivante est donc naturellement l'étude de la reproductibilité des indices de texture, en particulier dans l'objectif de réaliser des études multicentriques. Cette étape est d'autant plus critique que l'on peut supposer l'existence d'une influence réelle des performances et réglages d'acquisition des différents TDM sur la robustesse des indices présélectionnés. En effet, la grande majorité des études utilise des données de patients qui ont en commun des caractéristiques cliniques (même type de cancer, même type de protocole de traitement notamment), mais la provenance des données d'imagerie réalisées dans le suivi des patients est rarement contrôlée. Les patients « vont passer des examens » là où une place est disponible, avec des protocoles d'acquisition des images variables d'un centre à un autre, d'une machine à une autre, d'un médecin radiologue à un autre.

Dans les études réalisées, cette notion d'hétérogénéité des examens en terme de type/constructeur de machine mais également des paramètres pré-acquisition ne sont que rarement mentionnés et on considère souvent qu'un « scanner » est une donnée globale exploitable, sans discrimination. Cela avait du sens dans les études qui considéraient uniquement les données anatomiques de la maladie, comme les critères RECIST, car la localisation et la taille des lésions sont des données reproductibles. En revanche, si l'on souhaite étudier des paramètres de nature quantitative comme la « texture » de l'image de la tumeur, alors il devient impératif de s'assurer de la reproductibilité des valeurs des indices que l'on mesure. C'est donc dans ce contexte que j'ai entrepris cette étude au sein de mon service de radiologie puis étendu à l'échelle européenne au sein d'un consortium détaillé dans la partie méthodologie suivante.

3.1 Méthode

Cette étude a été menée au sein d'un consortium appelé Cancer Core Europe qui associe 7 centres européens de lutte contre le cancer qui ont décidé de s'allier à des fins de recherche[42,43] : Cambridge, Stockholm, Amsterdam, Barcelone, Heidelberg et Villejuif. Un des axes du projet portait sur l'imagerie médicale quantitative pour le cancer. A travers ce partenariat, différents types d'appareils de tomodensitométrie issus des différents centres ont été accessibles pour réaliser des expériences de reproductibilité avec le fantôme développé dans le cadre de ma thèse et utilisé dans l'étude précédente.

Dans la pratique, les acquisitions ont été réalisées sur six appareils (cf. tableau 4), issus de deux constructeurs différents (Siemens et GE) et résumées dans le tableau 5. Il s'agissait dans un premier temps de renouveler les mesures de répétabilité sur chaque appareil avec 8 acquisitions identiques (A à H) et compléter de 24 acquisitions supplémentaires pour lesquelles les paramètres d'acquisition étaient modifiés un à un. Cela a concerné la tension et intensité du tube à rayon X, l'épaisseur de coupe, le champ de vue, le pitch et l'index de reconstruction itérative. Les valeurs de ces paramètres d'acquisition ont été choisies pour être cohérentes avec celles utilisées en routine clinique.

Villejuif : GE Discovery CT750 HD (GE1)	Heidelberg *:* Siemens Somatom Definition Flash (Si1)
Villejuif : GE CT450 HD optima (GE2)	Cambridge: Siemens Somatom Definition AS+ (Si2)
Stockholm Karolinska : GE Revolution CT (GE3)	Barcelone: Siemens Sensation 64 (Si3)

Tableau 4 : Liste des références des six modèles de TDM mis en œuvre dans cette étude.

	Tube voltage (kVp)	Tube current (mAs)	Slice Thickness (mm)		filter		FOV^ (mm)	pitch	Iterative Reconstruct.	
			GE	Siemens	GE	Siemens			GE ASIR	Siemens SAFIRE
A	120	150	1.25	1	Standard	B40f	360	1	0	0
2	80	150	1.25	1	Standard	B40f	360	1	0	0
3	100	150	1.25	1	Standard	B40f	360	1	0	0
4	140	150	1.25	1	Standard	B40f	360	1	0	0
B	120	150	1.25	1	Standard	B40f	360	1	0	0
5	120	100	1.25	1	Standard	B40f	360	1	0	0
6	120	200	1.25	1	Standard	B40f	360	1	0	0
7	120	300	1.25	1	Standard	B40f	360	1	0	0
8	120	400	1.25	1	Standard	B40f	360	1	0	0
C	120	150	1.25	1	Standard	B40f	360	1	0	0
9	120	150	0.625	0.75	Standard	B40f	360	1	0	0
10	120	150	2.5	2	Standard	B40f	360	1	0	0
11	120	150	5	5	Standard	B40f	360	1	0	0
D	120	150	1.25	1	Standard	B40f	360	1	0	0
12	120	150	1.25	1	Standard	B30f	360	1	0	0
13	120	150	1.25	1	Soft	B35f	360	1	0	0
14	120	150	1.25	1	Bone	B75f	360	1	0	0
15	120	150	1.25	1	Lung	B80f	360	1	0	0
E	120	150	1.25	1	Standard	B40f	360	1	0	0
16	120	150	1.25	1	Standard	B40f	160	1	0	0
17	120	150	1.25	1	Standard	B40f	260	1	0	0
18	120	150	1.25	1	Standard	B40f	400	1	0	0
F	120	150	1.25	1	Standard	B40f	360	1	0	0
19	120	150	1.25	1	Standard	B40f	360	0.6		
20	120	150	1.25	1	Standard	B40f	360	1.3		
G	120	150	1.25	1	Standard	B40f	360	1	0	0
22	120	150	1.25	1	Standard	B40f	360	1	20	1
23	120	150	1.25	1	Standard	B40f	360	1	40	2
24	120	150	1.25	1	Standard	B40f	360	1	60	3
25	120	150	1.25	1	Standard	B40f	360	1	80	
26	120	150	1.25	1	Standard	B40f	360	1	100	
H	120	150	1.25	1	Standard	B40f	360	1	0	0

Tableau 5 : Résumé des paramètres d'acquisition des différentes expériences réalisées sur le fantôme sur les 6 différentes machines

Dans le cadre de notre étude sur la reproductibilité des indices de texture, nous avons considéré trois cas de figures :

1. Etude de la **répétabilité** des mesures

L'analyse faite ici est la même que précédemment décrite en déterminant le coefficient de variation (CV en %) des indices de texture à partir des 8 acquisitions de répétabilité.

2. Etude de la variabilité **inter-scanner** des mesures TDM :

Il s'agit de comparer la capacité des différents appareils TDM à mesurer les mêmes intensités et donc ainsi obtenir les mêmes valeurs d'indices de texture. Pour cela, nous avons déterminé :

- La valeur moyenne des indices de texture calculés à partir des 8 acquisitions de répétabilité ainsi que la valeur moyenne de l'intensité des pixels, cette dernière servant de point de comparaison entre les machines. La variation des mesures a été indiquée par le calcul de l'écart type (SD). En outre, la capacité des indices de texture à différencier deux milieux hétérogènes a été évaluée en relevant la différence des valeurs moyennes des caractéristiques de texture de C1 et C2.
- Un test non paramétrique de Kruskall-Wallis en vue d'établir si les mesures de chaque scanner peuvent être considérées comme similaires. Pour chacun des trois inserts et pour les neuf variables d'intérêt (valeur moyenne de l'intensité des pixels et les huit indices de texture retenu) obtenues à partir des 8 acquisitions de répétabilité, le test compare les valeurs ainsi obtenues à partir de toutes les combinaisons de scanners Siemens (S1 vs S2, S1 vs S3, S2 vs S3 et S1, S2, S3 ensemble), de même pour GE, et enfin tous les appareils ensemble. Une valeur p inférieure à 0,01 a été considérée comme statistiquement significative.

3. Etude de la variabilité **intra-scanner** des mesures en modifiant les paramètres d'acquisition

En plus de l'étude des coefficients de variation (CV) déjà définie précédemment, nous avons enrichi notre analyse en vue d'établir la reproductibilité des mesures dans un contexte d'études multicentriques avec la modification des paramètres d'acquisitions. Plus précisément, afin d'évaluer si la variation du calcul de l'indice de texture est liée à la modification du paramètre d'acquisition ou bien à la variabilité intrinsèque de la mesure, une analyse des ratios à été faite à partir des 24 autres acquisitions selon la définition suivante:

$$Ratio = \frac{SDvar}{SDint}$$

SDvar étant la déviation standard relative à la modification du paramètre d'acquisition (par exemple 4 mesures pour les 4 acquisitions avec 4 valeurs de kv différents) et SDint étant la déviation standard intrinsèque de la mesure répétée 8 fois dans les mêmes conditions d'acquisition.

Si le ratio est supérieur à 1, alors la variation de la mesure liée à la modification du paramètre d'acquisition est supérieure à la variation intrinsèque de la mesure. Nous avons considéré qu'un ratio supérieur à 5 était trop important pour une utilisation en routine clinique, en l'absence de recommandations sur ce sujet.

Enfin, l'analyse des trois régions d'intérêt a été réalisée selon la même méthodologie que dans l'étude précédente. Il est à noter que dans ce travail, la zone homogène est renommée C3 et les 2 zones hétérogènes identiques C1 et C2. A noter également que, s'agissant du scanner Discovery 750, nous avons réalisé un nouveau set d'acquisitions.

3.2 Résultats

3.2.1 Etude de la répétabilité

L'étude de répétabilité a été réalisée sur les différentes machines, à partir des huit acquisitions (nommées de A à H dans le tableau 5). Les résultats exprimés en coefficient de variabilité (%CV) sont résumés dans le tableau 7. Pour faciliter la lecture, les valeurs supérieures à 20% sont indiquées en jaune.

		Mean CT density	Skewness	Kurtosis	GLCM Contrast	Entropy	Dissimilarity	NGLDM Contrast	SZE	ZLNU
GE1	C1	0.7	-7.7	7.7	5.1	0.9	2.9	4.6	0.7	2.5
	C2	0.7	-1.4	3.0	4.3	0.6	2.2	4.9	2.3	6.6
	C3	5.0	209.5	4.3	6.4	3.0	4.5	31.9	10.7	22.8
GE2	C1	1.6	-29.8	27.9	16.0	0.9	2.2	10.9	1.7	6.1
	C2	2.1	-12.7	21.4	9.3	2.0	6.4	15.8	2.8	10.0
	C3	3.3	701.6	5.1	7.7	7.7	7.4	37.3	14.8	32.5
GE3	C1	6.3	-21.1	15.2	13.2	1.6	6.3	12.2	3.6	11.8
	C2	10.2	-17.8	19.2	24.1	1.3	8.8	10.1	1.8	7.7
	C3	22.7	121.3	9.8	14.8	5.1	8.2	16.7	7.6	26.7
Si1	C1	1.1	-4.1	7.0	15.7	1.2	8.4	5.3	6.1	6.4
	C2	1.1	-5.1	5.1	14.9	1.2	7.6	7.3	4.9	6.8
	C3	10.0	-69.2	18.5	32.1	11.0	25.6	33.1	18.5	21.9
Si2	C1	3.3	-7.2	1.6	1.5	0.4	0.7	5.2	0.9	2.5
	C2	1.2	-0.6	0.7	1.0	0.2	0.5	4.1	1.4	3.1
	C3	14.4	-47.9	16.3	7.7	3.8	5.1	35.5	10.0	26.1
Si3	C1	3.0	-4.0	2.0	2.3	0.3	1.3	3.8	1.1	4.2
	C2	4.3	-2.8	3.2	2.2	0.5	0.6	4.8	1.6	6.0
	C3	11.7	-100.5	7.1	7.2	2.7	5.6	21.7	9.3	12.5

Tableau 7 : Coefficients de variation observés pour 8 répétitions de la même acquisition sur 6 différentes machines, selon les 3 zones d'intérêt sélectionnées (C1 et 2 = zones hétérogènes, C3 = zone homogène). En jaune, les cv> 20%.

En ce qui concerne l'insert C3 homogène, les coefficients de variations sont globalement plus élevés que les l'ensemble des autres valeurs. Plus précisément, la valeur de cinq indices de texture présentait également une variabilité avec un CV inférieur à 20 % : HISTO_kurtosis, GLCM_contrast, GLCM_entropy, GLCM_dissimilarity et GLZLM_SZE. En revanche, HISTO_Skewness, NGLDM_Contrast, GLZLM_ZLNU ont des valeurs très dispersées jusqu'à des CV de 700% en raison de leurs définitions mathématiques souvent proches de 0.

Pour les inserts hétérogènes, C1 et C2, seuls deux indices de texture sont apparus avec des CV supérieurs à 20% : HISTO_Skewness pour seulement deux scans (GE2 et GE3) et HISTO_Kurtosis pour GE2 seulement. Toutes les autres caractéristiques avaient des CV inférieurs à 10%.

3.2.2 Variabilité inter-scanner de la mesure

Tout d'abord, avant d'analyser la variabilité des indices de texture, on s'intéresse à la variabilité du signal obtenu sur les images issues des six scanners en relevant la valeur moyenne des intensités des pixels en unité Hounsfield, et ce pour les 3 régions d'intérêt. Les valeurs sont reportées dans la première colonne du tableau 8, et leurs dispersions est illustrée sur la figure 11.

On constate que la valeur moyenne varie de 15% dans les régions d'intérêt hétérogènes mais varie jusqu'à 42% pour la région d'intérêt homogène. Comme attendu, les valeurs des intensités moyennes sont proches pour les 2 inserts hétérogènes, puisque leurs densités sont équivalentes et seul l'organisation des pixels de l'image diffère. On observe enfin l'absence de recouvrement des intensités moyennes entre l'insert homogène et les inserts hétérogènes.

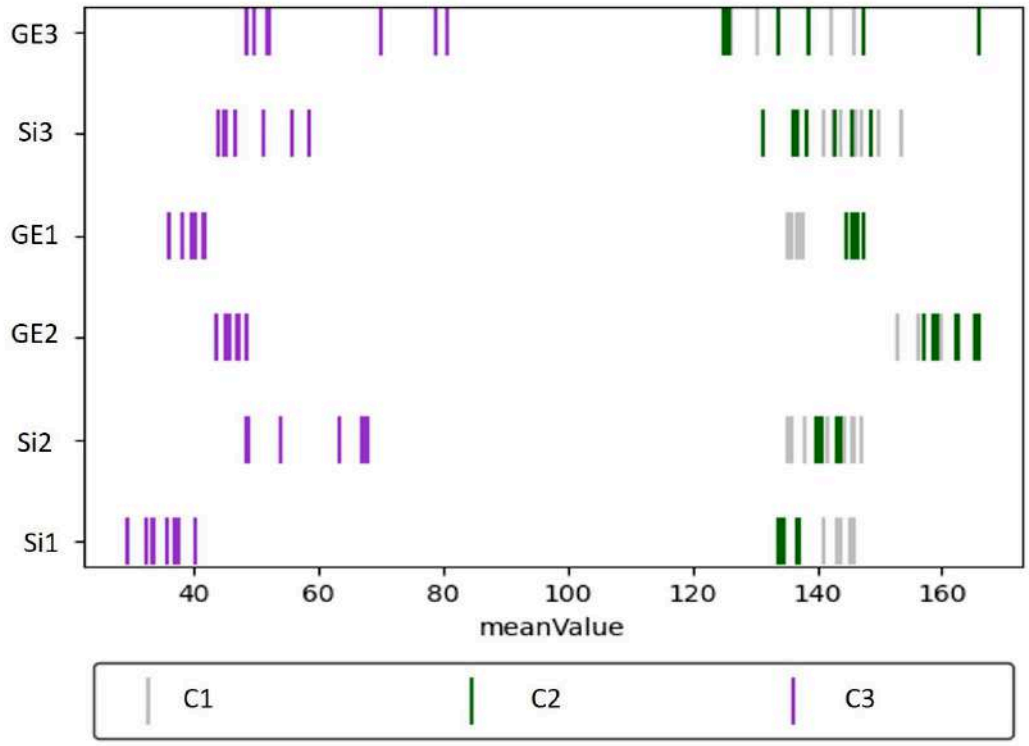


Figure 11 : représentation de la dispersion des valeurs moyennes des régions d'intérêt en unité Hounsfield : chaque bâtonnet représente la mesure sur 1 des 8 acquisitions identiques, respectivement pour chacune des machines (lignes) et chacune des zones d'intérêt (C1 en gris, C2 en vert et C3 en violet).

		Mean CT density (UH)	Skewness (AU)	Kurtosis (AU)	GLCM Contrast (AU)	Entropy (AU)	Dissimilarity (AU)	NGLDM Contrast (AU)	SZE (AU)	ZLNU (AU)
GE1	C1	136,3 ± 0,98	-1 ± 0,08	5,5 ± 0,42	28,1 ± 1,42	2,4 ± 0,02	3,8 ± 0,11	0,1 ± 0,01	0,8 ± 0,01	205,3 ± 5,16
	C2	145,8 ± 0,96	-2,2 ± 0,03	8,5 ± 0,25	70,8 ± 3,07	2,4 ± 0,01	5,2 ± 0,11	0,2 ± 0,01	0,8 ± 0,02	219,4 ± 14,48
	C3	39,3 ± 1,96	0,1 ± 0,14	3 ± 0,13	0,7 ± 0,04	0,9 ± 0,03	0,6 ± 0,03	0 ± 0,009	0,5 ± 0,05	14,9 ± 3,4
GE2	C1	156,4 ± 2,54	-1 ± 0,29	5,4 ± 1,51	26,7 ± 4,26	2,4 ± 0,02	3,6 ± 0,08	0,1 ± 0,01	0,7 ± 0,01	186,5 ± 11,31
	C2	162,1 ± 3,43	-1,7 ± 0,22	6,8 ± 1,46	50,8 ± 4,72	2,4 ± 0,05	4,8 ± 0,31	0,2 ± 0,03	0,8 ± 0,02	218,7 ± 21,92
	C3	46,2 ± 1,51	0 ± 0,1	2,9 ± 0,15	0,4 ± 0,03	0,7 ± 0,05	0,4 ± 0,03	0 ± 0,008	0,5 ± 0,07	11,4 ± 3,7
GE3	C1	133,6 ± 8,45	-1,2 ± 0,25	6 ± 0,92	23 ± 3,05	2,2 ± 0,04	3,4 ± 0,21	0,1 ± 0,01	0,7 ± 0,03	193,7 ± 22,82
	C2	137,2 ± 13,13	-1,9 ± 0,31	7,9 ± 1,31	49,1 ± 10,59	2,4 ± 0,03	4,6 ± 0,38	0,2 ± 0,02	0,8 ± 0,01	228,8 ± 13,47
	C3	60,4 ± 12,84	0,1 ± 0,08	3,1 ± 0,3	1,1 ± 0,15	1,1 ± 0,05	0,8 ± 0,06	0 ± 0,005	0,5 ± 0,04	60,4 ± 12,8
Si1	C1	144 ± 1,57	-1,2 ± 0,05	6,5 ± 0,45	33,2 ± 5,21	2,4 ± 0,03	4 ± 0,33	0,1 ± 0	0,6 ± 0,04	181,7 ± 11,54
	C2	135,4 ± 1,43	-1,3 ± 0,07	5,1 ± 0,26	67,7 ± 10,11	2,6 ± 0,03	5,7 ± 0,44	0,2 ± 0,01	0,7 ± 0,03	273,2 ± 18,5
	C3	34,8 ± 3,48	-0,1 ± 0,04	1,9 ± 0,36	1,4 ± 0,46	0,8 ± 0,09	0,9 ± 0,24	0,1 ± 0,021	0,4 ± 0,08	5,4 ± 1,17
Si2	C1	141,8 ± 4,7	-0,4 ± 0,03	4,9 ± 0,08	24,2 ± 0,37	2,4 ± 0,01	3,5 ± 0,02	0,1 ± 0	0,7 ± 0,01	195,5 ± 4,84
	C2	142,3 ± 1,64	-1,8 ± 0,01	6,8 ± 0,05	56,8 ± 0,59	2,4 ± 0,01	4,9 ± 0,02	0,2 ± 0,01	0,8 ± 0,01	207,9 ± 6,35
	C3	60,5 ± 8,71	-0,5 ± 0,26	3,5 ± 0,57	0,5 ± 0,04	0,8 ± 0,03	0,4 ± 0,02	0 ± 0,007	0,5 ± 0,05	7,7 ± 2
Si3	C1	146,2 ± 4,35	-0,7 ± 0,03	4,9 ± 0,1	34,5 ± 0,79	2,4 ± 0,01	4,2 ± 0,05	0,1 ± 0,004	0,8 ± 0,01	233,6 ± 9,7
	C2	140 ± 6,01	-1,6 ± 0,05	6,1 ± 0,19	63,9 ± 1,4	2,6 ± 0,01	5,6 ± 0,04	0,2 ± 0,01	0,8 ± 0,01	264,9 ± 15,95
	C3	49,4 ± 5,77	-0,2 ± 0,19	3 ± 0,21	0,6 ± 0,05	0,9 ± 0,02	0,5 ± 0,03	0 ± 0,005	0,5 ± 0,04	13,1 ± 1,64

Tableau 8 : Quantification des différentes valeurs : moyenne (mean) en unités Housfield, et des 8 indices de texture observées pour les 8 acquisitions identiques sur chacune des machines tomodensitométriques et pour les 3 régions d'interêt C1 C2 et C3.

Afin de tester si les **mesures réalisées sur des scanners différents pouvaient être considérées comme similaires**, nous avons appliqué le test non paramétrique de Kruskall Wallis. Les résultats sont résumés dans le tableau 9 en indiquant en couleur verte les valeurs p>0,01, c'est à dire l'absence de différence significative entre les mesures de la combinaison de scanner indiquée. Ainsi, il apparaît des mesures similaires d'indices de texture entre les trois scanners Siemens et distinctement deux scanners GE mais pas de tous les scanners ensemble.

	Mean	Skewness	Kurtosis	GLCM Contrast	Entropy	Dissimilarity	NGLDM Contrast	SZE	ZLNU
C1									
Si1 Si2	1.00	0.00	0.00	0.01	0.00	0.01	0.46	0.00	0.03
Si1 Si3	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.40	0.14
Si2 Si3	0.00	0.00	0.00	0.01	0.26	0.01	0.00	0.00	0.00
Si1 Si2 Si3	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GE1 GE2	0.00	0.00	0.40	0.40	0.06	0.25	0.00	0.09	0.07
GE2 GE3	0.11	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00
GE1 GE3	0.00	0.25	0.42	0.00	0.00	0.00	0.73	0.00	0.00
GE1 GE2 GE3	0.00	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00
all	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C2									
Si1 Si2	1.00	0.00	0.00	0.01	0.00	0.01	0.13	0.00	0.04
Si1 Si3	1.00	0.00	0.00	0.00	0.00	0.12	0.75	0.05	0.00
Si2 Si3	0.00	0.00	0.00	0.83	0.00	0.01	0.07	0.00	0.00
Si1 Si2 Si3	0.08	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00
GE1 GE2	0.00	0.96	0.40	0.00	0.39	0.40	0.43	0.25	0.29
GE2 GE3	0.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GE1 GE3	0.00	0.42	0.42	0.00	0.00	0.00	0.42	0.00	0.00
GE1 GE2 GE3	0.00	0.05	0.06	0.00	0.00	0.00	0.03	0.00	0.00
all	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
C3									
Si1 Si2	0.00	0.03	0.00	0.01	0.60	0.01	0.00	0.21	0.00
Si1 Si3	0.00	0.00	0.21	0.00	0.11	0.00	0.09	0.40	0.01
Si2 Si3	0.01	0.05	0.00	0.01	0.00	0.01	0.00	0.06	0.00
Si1 Si2 Si3	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.14	0.00
GE1 GE2	0.00	0.00	0.02	0.00	0.00	0.00	1.00	0.92	0.02
GE2 GE3	0.02	0.00	0.13	0.00	0.00	0.00	0.06	1.00	0.00
GE1 GE3	0.64	0.04	0.16	0.00	0.00	0.00	0.25	0.82	0.06
GE1 GE2 GE3	0.00	0.00	0.03	0.00	0.00	0.00	0.21	0.99	0.00
all	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.38	0.00

Tableau 9 : Tests de Kruskall Wallis appliqué aux différents échantillons de mesure. Les p-valeurs supérieures à 0.01 (en vert) suggèrent que l'hypothèse de similarité de la combinaison considérée ne peut pas être rejetée.

On s'intéresse aux indices de texture dont **les valeurs ne varient pas de plus de 20%** sur l'ensemble des machines, pour les 2 zones hétérogènes. 4 indices (GLCM_entropy, GLCM_dissimilarity, GLZLM_SZE, and GLZLM_ZLNU) ont cette caractéristique, mais le test de Kruskall Wallis ne retient que GLZLM-SZE pour l'analyse de similarité inter-machines. Toutefois, lorsque l'on considère la représentation de la dispersion des valeurs de cet indice en fonction des machines on observe un chevauchement important des valeurs mesurées en C1 et C2 (figure 12) qui implique par conséquence un test statistique positif à la similarité des mesures.

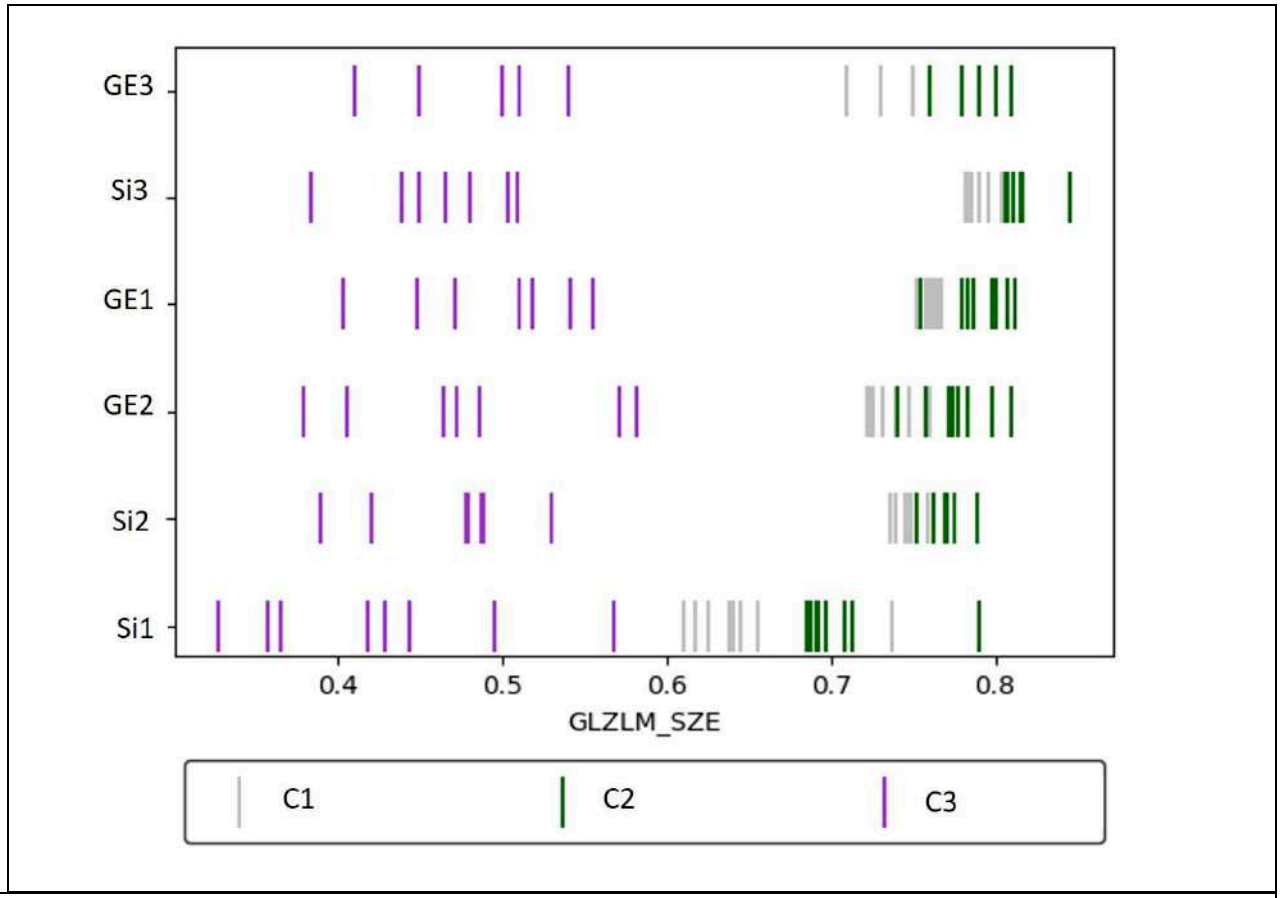


Figure 12 : représentation de la dispersion des valeurs de l'indice GLZLM-SZE des 3 régions d'intérêt: chaque bâtonnet représente la mesure sur 1 des 8 acquisitions identiques, respectivement pour chacune des machines (lignes) et chacune des zones d'intérêt (C1 en gris, C2 en vert et C3 en violet). Il existe de nombreuses zones de chevauchement sur les 2 régions hétérogènes alors que l'on souhaiterait pouvoir discriminer les zones l'une de l'autre grâce à la texture.

De plus, si l'on considère les indices de texture qui **permettent de discriminer les 2 régions hétérogènes C1 et C2 l'une de l'autre**, on constate que 5 indices ont cette propriété : HISTO_skewness, HISTO_kurtosis, GLCM_contrast GLCM_dissimilarity and NGLDM_Contrast. Ainsi, **par éliminations successives** des indices non reproductibles et de ceux qui ne permettent pas de différencier 2 zones hétérogènes dont les pixels constitutifs sont proches mais l'organisation spatiale est différente, un seul indice est retenu : il s'agit de GLCM_Dissimilarity (figure 13).

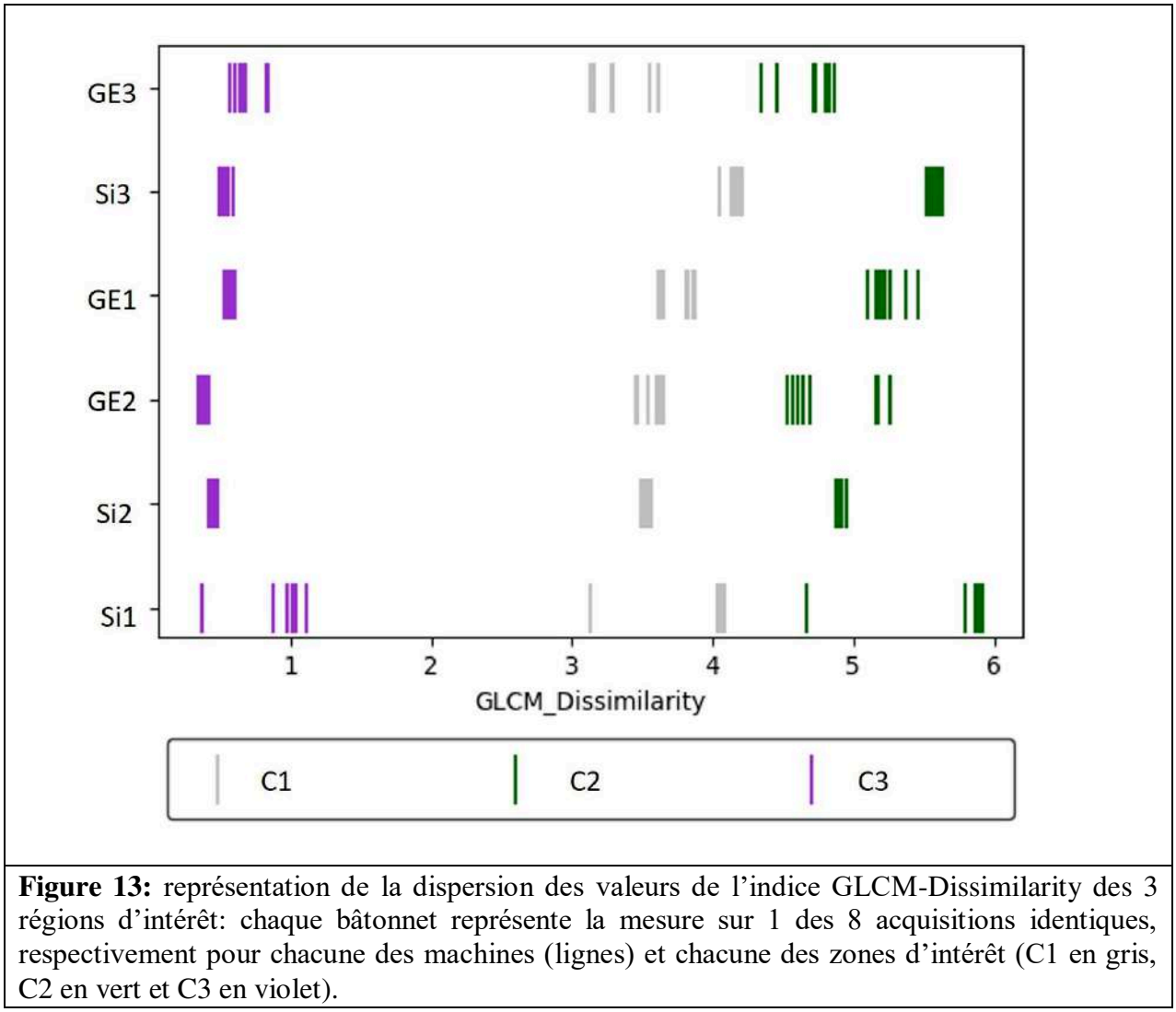


Figure 13: représentation de la dispersion des valeurs de l'indice GLCM-Dissimilarity des 3 régions d'intérêt: chaque bâtonnet représente la mesure sur 1 des 8 acquisitions identiques, respectivement pour chacune des machines (lignes) et chacune des zones d'intérêt (C1 en gris, C2 en vert et C3 en violet).

3.2.3 Variabilité intra scanner selon les modifications des paramètres d'acquisition

L'étape suivante consiste à analyser l'influence respective des paramètres d'acquisition sur le calcul des indices de texture.

Nous avons ainsi analysé les 8 indices de texture, calculés sur 3 régions d'intérêt (C1 C2 et C3) d'images obtenues sur 7 sets différents d'acquisitions scanographiques (modifications respectives de 7 paramètres d'acquisition parmi tension du tube en kV (4 valeurs), intensité du tube en mAs (5 valeurs), épaisseur de coupe en mm (4 valeurs), pitch (3 valeurs), champs de vue en mm (4 valeurs), indice de reconstruction itérative(6/4 valeurs), filtre(5 valeurs)) sur 6 machines différentes ; soit un total de 3 (régions d'intérêt) x 8 (indices de texture) x x 6 (machines) = 144 ; 144 x 7 (paramètres d'acquisition) = 1008 calculs de ratios.

Le tableau 10 décompte le nombre de ratio > 5 et donc considérés comme trop variables.

	kV ratio	mA ratio	ST ratio	pitch ratio	FOV ratio	IR ratio	filter ratio
GE1	3	0	13	1	8	3	16
GE2	2	1	2	0	4	0	13
GE3	0	0	0	NA	3	0	15
Si1	2	0	0	0	3	0	16
Si2	8	0	11	0	9	0	21
Si3	6	2	13	0	10	NA	21
TOTAL	**21/144**	**3/144**	**39/144**	**1/144**	**37/144**	**3/144**	**102/144**

Tableau 10 : Résumé du nombre de ratio supérieur à 5 lorsque l'on modifie les paramètres d'acquisition (colonnes) et selon la machine considérée (lignes)

Au total, 20% (206/1008) des mesures sont modifiées de manière importante (ratio>5) lorsque l'on modifie les paramètres d'acquisition. Avec 102 ratios >5, le filtre est le principal facteur qui influence la variabilité, et ce pour les 8 indices de texture analysés.

La modification d'autres paramètres d'acquisition a moins d'influence sur la variabilité de la mesure : intensité et tension du tube, pitch et index de reconstruction itérative.

La modification de l'épaisseur de coupe et du champ de vue altère également la mesure, mais il s'agit de paramètres que l'on peut maitriser au moment de l'acquisition.

Afin d'illustrer le type d'écarts constatés, le tableau 11 montre une partie de l'analyse sur une seule machine (Si2)

C1 Si2	kV ratio	mA ratio	Ep coupe ratio	pitch ratio	Champs de vue ratio	SAFIRE ratio	filtre ratio
HISTO_Skewness	11.7	0.7	8.2	2.2	1.5	0.9	23.8
HISTO_Kurtosis	1.3	1.1	6.8	0.3	1.5	0.9	24.6
GLCM_Entropy_log10	2.4	0.8	12.0	1.3	7.4	0.6	31.2
GLCM_Dissimilarity	6.7	1.7	23.2	1.9	48.4	0.6	196.9
GLCM_Contrast	5.4	1.4	17.6	1.3	34.1	0.3	390.4
NGLDM_Contrast	3.0	1.0	2.3	0.1	11.9	0.7	49.3
GLZLM_SZE	2.4	1.0	5.5	1.1	12.9	2.9	13.0
GLZLM_ZLNU	2.6	0.6	5.9	1.2	23.2	2.1	24.3
C2 Si2	kV ratio	mA ratio	Ep coupe ratio	pitch ratio	Champs de vue ratio	SAFIRE ratio	filtre ratio
HISTO_Skewness	15.3	1.8	49.7	1.5	5.6	0.6	46.8
HISTO_Kurtosis	9.9	1.8	37.5	0.5	7.2	1.0	59.9
GLCM_Entropy_log10	7.8	1.1	1.7	2.0	5.0	1.1	28.1
GLCM_Dissimilarity	2.7	1.1	26.5	2.0	71.0	1.8	230.2
GLCM_Contrast	3.4	1.4	30.1	1.5	53.3	1.8	476.7
NGLDM_Contrast	1.8	0.8	2.7	1.7	14.8	1.2	42.3
GLZLM_SZE	0.7	1.8	2.4	0.6	4.5	1.5	5.9
GLZLM_ZLNU	0.7	2.1	3.4	0.7	31.1	1.9	13.8
C3 Si2	kV ratio	mA ratio	Ep coupe ratio	pitch ratio	Champs de vue ratio	SAFIRE ratio	filtre ratio
HISTO_Skewness	1.2	1.5	0.8	0.9	0.6	0.4	1.3
HISTO_Kurtosis	0.8	0.8	0.6	0.7	0.1	0.7	0.4
GLCM_Entropy_log10	7.0	3.5	4.6	2.0	0.9	1.5	20.6
GLCM_Dissimilarity	7.5	4.5	3.9	3.0	2.3	2.3	52.1
GLCM_Contrast	8.6	3.7	3.0	2.5	1.8	1.8	161.9
NGLDM_Contrast	0.8	1.0	0.7	1.2	1.0	1.0	6.2
GLZLM_SZE	0.7	2.5	0.8	1.0	2.8	1.2	2.8
GLZLM_ZLNU	2.3	2.0	1.1	0.8	0.2	1.6	42.2

Tableau 11: Ratio calculés sur la machine Si2 et illustrant la variation liée à la modification du parametre d'acquisition relativement à la déviation standard de la mesure réalisée sur 8 acquisitions identiques, respectivement pour les 3 régions d'intérêt C1 C2 et C3. Les ratio > 5 sont surlignés en jaune.

3.3 Discussion

Les recherches actuelles dans le nouveau champ d'application de la Radiomique en imagerie médicale, à la fois pour une caractérisation tissulaire et pour l'analyse de la réponse tumorale sont fondées sur le principe de modifications « invisibles à l'œil nu », et que l'on pourrait mettre en évidence par l'application de matrices mathématiques sur une région d'intérêt traduite en valeurs mathématiques (chaque pixel d'une image ayant une valeur numérique contenue dans l'image).

Cette approche est très séduisante puisqu'elle permettrait d'examiner de manière beaucoup plus précise les lésions que nous rencontrons quotidiennement et permettrait d'expliquer des phénomènes que nous ne comprenons pas totalement, en oncologie notamment. Par ailleurs, l'ambition de tendre vers une imagerie quantitative est un élément essentiel pour partager et discuter nos résultats à l'échelle internationale et conduire des études multicentriques.

Dans ce travail de thèse, nous avons quantifié la variabilité de la mesure des indices de texture sur le même objet physique, sur plusieurs types de machines scanographiques et en faisant varier de manière contrôlée les paramètres d'acquisition. Cette méthodologie qui s'est voulue des plus rigoureuses nous permet de mettre en évidence plusieurs éléments importants.

Tout d'abord, la variabilité du signal en unité Hounsfield entre les scanners qui imagent le même objet, ce qui est particulièrement surprenant au regard des nombreuses mesures qualités faites régulièrement dans nos services de radiologie. L'étude de répétabilité a montré également une faible variabilité (coefficient de variation < 20%) sur l'ensemble des six scanners pour 5 indices de texture : GLCM_contrast, GLCM_entropy, GLCM_dissimilarity, GLZLM_SZE and GLZLM_ZLNU.
Ainsi, ces 5 indices pourraient servir de biomarqueurs potentiels pour caractériser une lésion au moment du diagnostic, du point de vue de son hétérogénéité.

Ensuite, l'étude de la variabilité inter-scanner a mis en évidence 4 indices dont les coefficients de variation sont inférieurs à 20% et qui seraient donc robustes pour une utilisation multicentrique. Il s'agit de SZE, GLCM_entropy, GLCM_dissimilarity, and GLZLM_ZLNU. Néanmoins, s'agissant de SZE, bien que l'analyse statistique montre une bonne reproductibilité, celle-ci est en réalité une construction statistique liée à la grande dispersion de ses valeurs et nous pensons qu'il ne faudrait pas retenir cet indice pour des analyses multicentriques.

Enfin, il est important qu'un indice de texture soit capable de discriminer une zone homogène d'une zone hétérogène, et 2 zones hétérogènes, proches en valeurs de densité moyenne, mais différentes visuellement. Notre étude permet de sélectionner pour cette caractéristique discriminante 5 indices de texture : HISTO_skewness, HISTO_kurtosis, GLCM_contrast, GLCM_dissimilarity et NGLDM_Contrast. Ces indices pourraient donc servir de biomarqueurs potentiels dans l'identification de modifications au sein d'une lésion sous traitement, sous réserve que l'analyse soit réalisée dans les mêmes conditions, c'est à dire sur la même machine, avec les mêmes paramètres d'acquisition.

Au total, dans nos conditions expérimentales, GLCM_dissimilarity apparaît être le seul indice de texture qui a, à la fois, une faible variabilité inter-scanner et un bon pouvoir discriminant. Ces

travaux qui sont en cours de publication au sein du consortium Cancer Core Europe, nous questionnent sur ma méthode pour améliorer la fiabilité de la mesure des indices de texture en tomodensitométrie.

Nous pouvons faire l'hypothèse forte que la variabilité inter-scanner, sur le même objet, et sur 8 acquisitions identiques réalisées au mieux avec les mêmes paramètres d'acquisition sont liées aux performances intrinsèques de chaque scanner. Cette hypothèse est égaleemnt partagée par d'autres équipes dont celle de Ger et al, qui a conduit des mesures sur 100 machines différentes. Leur méthodologie est un peu différente de la nôtre, mais les auteurs montrent qu'en fixant le protocole d'acquisition, la variabilité de la mesure des indices de texture est améliorée. Notre étude permet de préciser ce point en montrant de manière plus fine quels paramètres d'acquisition sont responsables de la variabilité et il ressort que le filtre appliqué est l'un des plus importants, tandis que l'intensité du courant, le pitch ou le degré de reconstruction itérative ont peu d'influence sur la mesure. On pourrait donc envisager de proposer un **"protocole d'acquisition dédié à la radiomique en clinique",** qui prendrait en compte le besoin de fiabilité de la mesure des indices tout en garantissant la radioprotection des patients. Ceci est limité par la grande variabilité du signal en unité Hounsfield issue de la diversité des constructeurs et des technologies embarquées dans nos machines scanographiques. Il parait donc majeur au vu de nos résultats (sur 6 machines seulement) et des résultats de Ger et al, d'alerter les constructeurs de scanners sur leur rôle dans le développement d'une vraie discipline radiomique, et de la nécessité de standardiser au moins partiellement une partie de leur technologie à des fins de recherche. Ce dernier point fait l'objet d'un consortium international appelé QIBA pour Quantitative Imaging Biomakers Alliance.

Au vu de nos résultats, nous comprenons que la Radiomique en tant que discipline d'analyse « avancée » des images médicales doit préalablement s'astreindre à un effort méthodologique visant à améliorer la reproductibilité et la fiabilité des données extraites des images.
Les différentes études précédemment citées plaident toutes pour cet effort d'harmonisation : il est impératif d'aller vers des acquisitions a priori avant d'envisager d'utiliser les résultats dans le cadre d'études mono ou multicentriques en tant que biomarqueurs. Nous comprenons aussi que le rôle et l'implication des constructeurs est fondamental.

Enfin, dans l'hypothèse où un consensus réel serait possible au sein des différentes parties prenantes (constructeurs et chercheurs en radiomique), une autre difficulté devra être prise en compte, celle de la transposabilité en routine clinique. L'influence du positionnement du patient lors de l'acquisition des images, de son morphotype et de l'injection ou non de produit de contraste vont également être importants à étudier. Il y a donc encore un long chemin à parcourir avant de pouvoir crier victoire sur le développement de la radiomique comme biomarqueur fiable en imagerie.

Publication (en cours de révision) dans **European Radiology** : Towards a standardization of radiomic studies on CT images: a multicenter phantom Study within the Cancer Core Europe consortium
Caramella C, Bach M, Cournede PH, Jaranowski P, Beets-Tan RGH, Gallagher FA, Sala E, Woitek R, Perez-Lopez R, Blomqvist L, Holmin S, Giandini T, Vaiani M, Messina A, Kleesiek J, Murray J, Prinz S, Sedlaczek O, Schlemmer HP, Oberrauch P, Lassau N, Pitre-Champagnat S.
for the CCE – Imaging Task Force

Conclusion 1

Dans ce chapitre, je me suis focalisée à mettre en place une étude méthodologique la plus robuste possible pour répondre à des questions soulevées par l'apparition de la radiomique appliquée à l'imagerie médicale et de son ambition d'extraire des images des valeurs quantitatives. Mon approche est tout à fait valable sur un objet physique pour des études de répétabilité.
Mais qu'en est-il pour les analyses sur des centaines de patients, avec tout autant d'indices d'imagerie extrait ? C'est ici que la frontière avec les méthodes mathématiques d'intelligence artificielle apparaît. En effet, le développement d'analyse statistiques puissantes via des réseaux de neurones a commencé à apparaitre dans le domaine médical et particulièrement en imagerie, avec des débuts très prometteurs et des déclarations tonitruantes telles que celles de Geoffrey Hinton en 2017 dans le NewYorker : « *if you work as a radiologist, you're like the coyote that's already over the edge of the cliff, but hasn't yet looked down so doesn't realise there's no ground underneath him. People should stop training radiologists now. It's just completely obvious that within 5 years, deep learning is going to do better than radiologists* »[44].

Des avancées incontestables ont été permises grâce à ces techniques, mais là aussi, un long chemin reste à parcourir en radiologie[45]. Il ne s'agit pas de rejeter le progrès technologique, ce qui serait impossible, mais d'accompagner ces avancées avec un raisonnement scientifique rigoureux[46]. Les sociétés savantes en sont d'ailleurs bien conscientes[47].

Nous verrons dans le prochain chapitre comment d'autres bouleversements récents dans le domaine médical et particulièrement en oncologie, m'ont donné des nouvelles perspectives d'études. Nous aborderons ainsi le thème de l'immunothérapie et de la place de l'imagerie dans l'analyse des réponses à ce nouveau type de traitement, puis le thème des anomalies génomiques au sein des tissus tumoraux, en particulier dans le cancer bronchique, et la place que l'imagerie peut jouer dans le parcours du patient.

Chapitre 3: Une autre approche de l'évaluation tumorale : le modèle de l'immunothérapie

1/ Contexte

Parallèlement aux espoirs naissants autour de l'utilisation de la Radiomique comme nouveau biomarqueur en oncologie, de nouveaux types de traitements innovants ont été développés et administrés à des patients dont les pronostics de survie étaient catastrophiques.

Au début des années 2010, l'apparition des immunothérapies par inhibition des check points moléculaires a ainsi totalement bouleversé la prise en charge des patients en oncologie, notamment chez ceux qui recevaient ce type de traitement au stade métastatique de leur maladie. De nombreuses publications enthousiastes ont montré un vrai bénéfice sur la survie des malades, et notamment l'obtention de réponses complètes et prolongées y compris après l'arrêt du traitement[48]. Ce bouleversement a été couronné à 2018 par l'attribution du Prix Nobel de Médecine à l'américain James P. Allison et au japonais Tasuku Honjo pour leurs « découverte d'un traitement contre le cancer par inhibition de la régulation immunitaire négative (« for their discovery of cancer therapy by inhibition of negative immune regulation »).

Comme tout nouveau traitement, la réponse tumorale est évaluée à travers des bilans d'imagerie réguliers avec la recherche de signe de progression qui font stopper le traitement pour manque d'efficacité. Pour cela, les critères d'imagerie habituellement utilisés sont les critères RECIST 1.1[1]. Néanmoins, de nouveaux types de réponses inhabituels sont apparus avec ces traitements.

Pseudoprogression

Au début de l'utilisation de ces nouvelles molécules, chez des patients atteints de mélanome métastatique, des cliniciens ont rapporté des cas de « pseudo-progressions ». Il s'agissait de patients dont la maladie progressait de manière indéniable en imagerie, qui stoppaient tout traitement, et qui, de manière tout à fait inattendue, revenaient en consultation avec des examens montrant une réponse objective[49]. Cet évènement, inédit jusqu'alors, a ensuite été décrit de plus en plus fréquemment (jusqu'à 9% des patients suivis pour mélanome), pour aboutir à des recommandations particulières aux patients qui reçoivent ce type de traitement : ne pas se fier à une imagerie qui montre une progression selon les critères habituels RECIST 1.1 et attendre entre 4 et 8 semaines une confirmation de la progression (ou le constat d'une réponse)[14].

Les critères RECIST 1.1 ont donc été mis en défaut et ont été adaptés avec la publication de

nouveaux guidelines pour les essais thérapeutiques, les critères iRECIST, dont l'objectif était principalement de permettre une collection homogène des données.

iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. Seymour L, Bogaerts J, Perrone A, Ford R, Schwartz LH, Mandrekar S, Lin NU, Litière S, Dancey J, Chen A, Hodi FS, Therasse P, Hoekstra OS, Shankar LK, Wolchok JD, Ballinger M, **Caramella C**, de Vries EGE; RECIST working group. **Lancet Oncol. 2017 Mar**[50]

Hyperprogression

A côté des patients qui présentent une « fausse » progression de la maladie, plusieurs équipes de cliniciens ont également constaté que la maladie de certains autres patients paraissait accélérée au début du traitement par immunothérapie. Ce phénomène également inattendu, a été appelé hyperprogression. L'hyperprogression pourrait en partie expliquer l'aspect des courbes de survie rapportées dans les différents essais thérapeutiques, avec un croisement des courbes en début de traitement, qui indique une « surmortalité » initiale des malades sous immunothérapies comparativement au groupe témoin sous chimiothérapie (figure 14)[51].

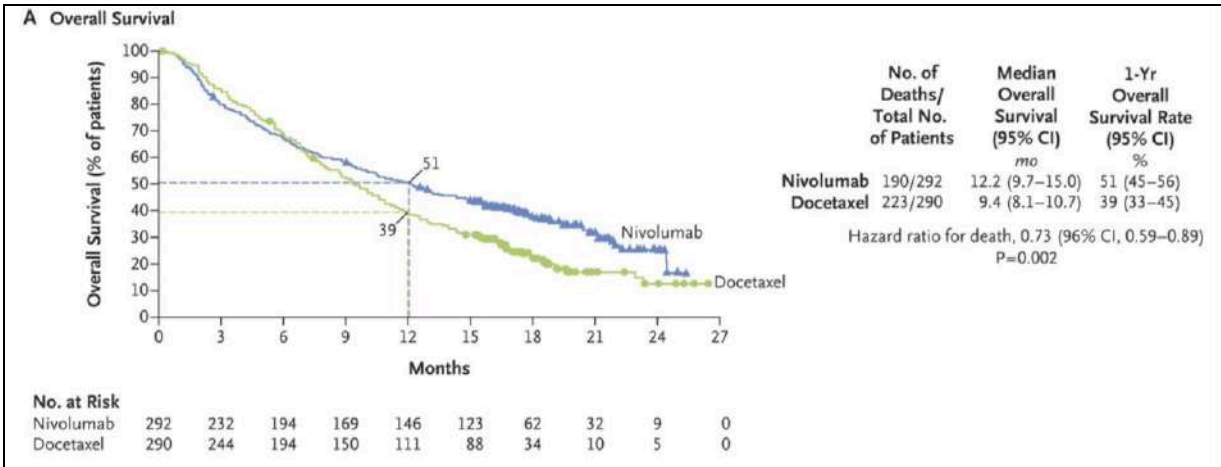
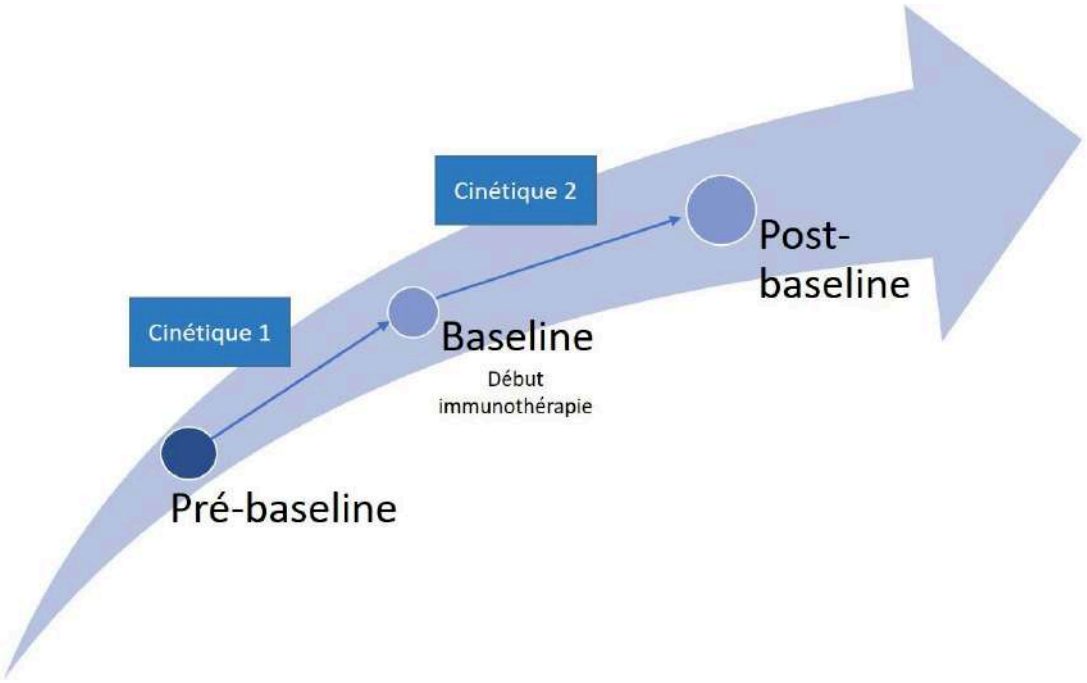


Figure 14 : from Borghaei H et al[51], courbe de survie de Kaplan Meyer de la première étude contrôlée randomisée évaluant le Nivolumab (immunothérapie) en comparaison au Docetaxel (traitement de référence) chez les patients ayant un cancer du poumon non à petites cellules. On constate un croisement des courbes en début de traitement indiquant une surmortalité chez les patients sous Nivolmab, qui pourrait être partiellement expliquée par le phénomène d'hyperprogression.

Afin d'évaluer l'incidence du phénomène, il est nécessaire de prendre en compte la cinétique tumorale, ce qui n'est pas le cas avec les critères d'évaluation RECIST 1.1, qui ne considèrent pas l'intervalle de temps entre les différentes évaluations.

Pour étudier l'hypothèse d'une accélération de la cinétique tumorale, on doit comparer 2 étapes : l'évolution de la maladie avant le début de l'immunothérapie et l'évolution de la maladie

après le début de l'immunothérapie. Pour cela, on a besoin de 3 imageries : une imagerie pré-baseline, l'imagerie baseline, et l'imagerie post-baseline (préBL, BL et postBL) :



Pour l'évaluation de la cinétique, la méthode choisie par plusieurs équipes internationales repose sur le Tumor Growth Rate (TGR), qui utilise les principes de RECIST, à savoir la détermination de lésions cibles mesurables représentatives de la maladie. Ces lésions doivent exister lors des 3 évaluations, ce qui a pour conséquence la non-prise en compte de l'apparition éventuelle de nouvelles lésions[52].

On va ensuite comparer la cinétique évolutive entre les points [préBL et BL] versus [BL et postBL] qui vont définir un TGR pre treatment et un TGR on treatment, et un ΔTGR qui est la différence entre les 2 TGR. Le TGR inclut une modélisation mathématique de la cinétique « théorique » exponentielle de la maladie et intègre la notion de temps (intervalle entre les évaluations successives qui diffère d'un patient à l'autre).

Les maladies étant en principe en progression avant le changement de traitement (TGRpré positif), on va obtenir 3 cas (figure 15) :

- ΔTGR négatif lorsque la maladie a une cinétique plus lente après la mise sous traitement qu'avant (ie le traitement est efficace sur la cinétique tumorale).
- ΔTGR nul quand la cinétique reste la même (ie : le traitement ne change rien à la cinétique).
- ΔTGR positif lorsque la maladie est accélérée après le début du nouveau traitement.

C'est dans ce 3eme cas de figure que l'on va s'interroger sur le lien de causalité entre l'introduction de ce nouveau traitement et l'accélération tumorale.

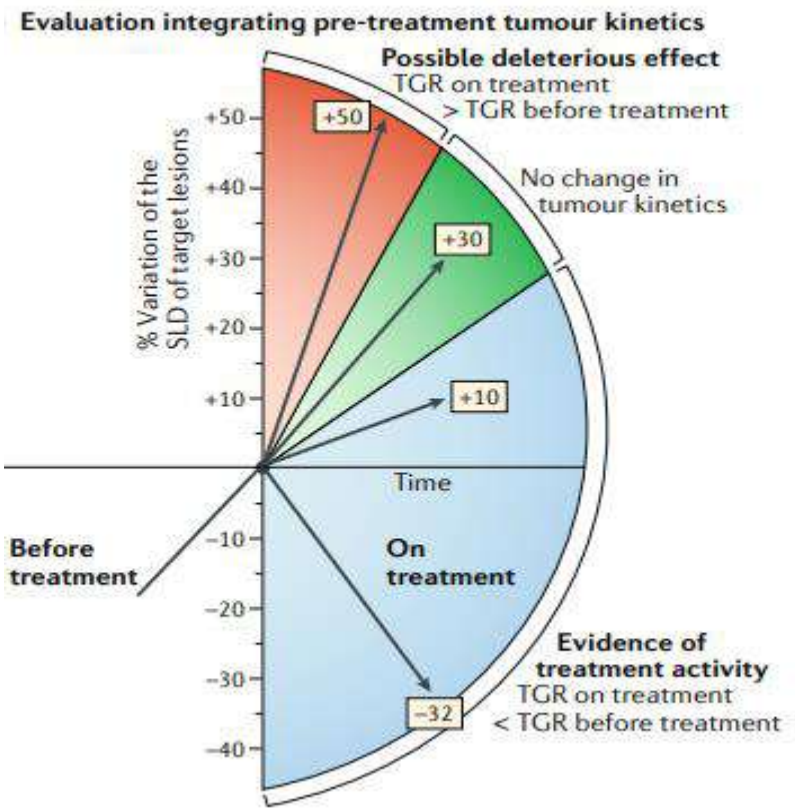


Figure 15 : illustration tirée de Champiat et al [53]. « The assessment of tumour kinetics involving the integration of the time between tumour evaluations and of data from a pretreatment assessment enables the dynamic and quantitative evaluation of the effects of treatment. CR, complete response; PR, partial response; TGR, tumour growth rate »

2/ Hyperprogression chez les patients sous immunothérapie pour un cancer bronchique à petites cellules

La constatation de ce nouveau phénomène clinique a été et reste débattu : certaines équipes pensaient qu'il s'agissait d'une évolution « normale » d'une maladie en poussée évolutive, tandis que d'autres pensaient qu'il pouvait s'agir d'une accélération de la cinétique tumorale liée à la mise en place du traitement. Nous avons donc réalisé une étude rétrospective afin d'évaluer l'incidence de l'hyperprogression dans une population de patients traités par immunothérapie et dans une population témoin.

Matériel et Méthode

Il s'agit d'une étude multicentrique rétrospective portant sur 406 patients recevant une immunothérapie pour un cancer bronchique métastatique et ayant eu 3 examens scannographiques disponibles (prébaseline, baseline de l'immunothérapie et postbaseline).

Une cohorte "contrôle" de patients n'ayant pas reçu d'immunothérapie mais une chimiothérapie classique a été évaluée selon les mêmes critères.

Nous avons relu l'ensemble de l'imagerie en aveugle du traitement reçu, afin d'identifier des lésions mesurables (selon les critères RECIST 1.1) présentes sur les 3 évaluations pour chaque patient.
Les TGR-pré et -on ont été calculé suivant la définition de Ferté et al [54]:

« Assuming tumor growth follows an exponential law, Vt, tumor volume at time t (expressed in months in the tumor evaluation) is equal to Vt = V0 exp (TG.t), where V0 is volume at baseline, and TG is the growth rate. We approximated the tumor volume (V) by V=4 π R3/3, where R, the radius of the sphere, is equal to D/2. Consequently, TG is equal to TG=3 log(Dt/D0)/(t1-t0), where log represents natural logarithms and t1-t0 is the time between evaluations in months. TGR results were reported as a percent increase in tumor volume per month using the following transformation: TGR=100 [exp(TG)−1], where exp(TG) represents the exponential of TG1 . »

Le ΔTGR était calculé sur la différence entre TGRpost et TGRpré. En l'absence de définition consensuelle, nous avons considéré qu'un ΔTGR >50% définissait l'hyperprogression.

Les caractéristiques des 2 cohortes de patients sont résumées dans les tableaux 12 et 13

Variable	No./Total No. (%)			Fisher Exact Test *P* Value
	Total (N = 406)	Non-HPD (n = 350)	HPD (n = 56)	
Age, y				
≥65	188 (46.3)	166 (47.4)	22 (39.3)	.31
<65	218 (53.7)	184 (52.6)	34 (60.7)	
Smoking history				
Current/former	371 (91.4)	319 (91.1)	52 (92.9)	>.99
None	35 (8.6)	31 (8.9)	4 (7.1)	
Smoking exposure, pack-years				
≤30	136 (33.5)	115/312 (36.9)	21/50 (42.0)	.53
>30	226 (55.7)	197/312 (63.1)	29/50 (58.0)	
Missing	44 (10.8)	38	6	
Histology				
Nonsquamous	294 (72.4)	252 (72.0)	42 (75.0)	.75
Squamous	112 (27.6)	98 (28.0)	14 (25.0)	
Stage[a]				
III	70 (17.2)	61 (17.4)	9 (16.1)	>.99
IV	336 (82.8)	289 (82.6)	47 (83.9)	

Tableau 12 : Caractéristiques des patients traités par immunothérapie selon l'âge, les antécédent de tabagisme, , l'histologie et le stade de la tumeur.

	N=59 No. (%)
Age	
≥ 65 years	28 (47%)
Smoking status	
Never smoker	6 (10%)
Former/current	53 (90%)
Histology	
Non-squamous	47 (80%)
Stage [a]	
III	7 (12%)
IV	52 (88%)

Tableau 13: Caractéristiques des patients dtraités par chimiothérapie (groupe témoin) selon l'âge, les antécédent de tabagisme, , l'histologie et le stade de la tumeur

L'évolution des cinétiques tumorales par le ΔTGR est résumée dans le tableau 14

	PD-1/PD-L1 inhibitors (N=406) No. (%)	Single-agent chemotherapy (N=59) No. (%)
TGR pre-treatment		
≤ 0	75 (18%)	14 (24%)
> 0	331 (82%)	45 (76%)
TGR on-treatment		
≤ 0	167 (41%)	39 (66%)
> 0	239 (59%)	20 (34%)
***delta* TGR[a]**		
≤ 0%	266 (66%)	47 (80%)
>0% and ≤50%	78 (19%)	9 (15%)
> 50%	62 (15%)	3 (5%)

a TGR on treatment minus TGR pre-treatment

Tableau 14: Résumé des comportements tumoraux avant la mise sous traitement (TGR prétreatment) et sous traitement (TGR on treatment), et résultante (delta TGR) dans les 2 groupes de patients sous immunothérapie et sous chimiothérapie.

Cette étude montre ainsi qu'il existe une différence de comportement tumoral à l'introduction du traitement par immunothérapie, avec chez 62 patients (15%) l'identification d'une hyperprogression, alors que dans la cohorte contrôle, seulement 3 cas (5%) répondaient à la définition d'hyperprogression.
Cette étude était la première comportant une population témoin.

Elle a également suggéré que l'hyperprogression survient plus fréquemment chez des patients avec un nombre d'organes touchés par des métastases supérieur (mais il n'y avait pas de signification statistique sur d'autres caractéristiques telles que l'âge ou le statut mutationnel).
Plusieurs points sont à souligner:

- le caractère retrospectif de l'étude
- l'exclusion des patients qui n'avaient pas de scanner post baseline (n=33) qui pourrait suggérer une sous-évaluation du phénomène d'hyperprogression
- l'exclusion des lésions qui n'étaient pas présentes lors des 3 évaluations : le TGR est construit sur l'analyse des mêmes lésions sur les 3 évaluations préBL, BL et post BL et ne prend donc pas en compte l'éventuelle apparition de nouvelle lésion au BL ou au post BL.
- la comparaison à d'autres études sur la cinétique tumorale et l'hyperprogression est rendue difficile par l'absence de définition consensuelle de l'hyperprogression.

Publication dans **Jama Oncol .** Hyperprogressive Disease in Patients With Advanced Non-Small Cell Lung Cancer Treated With PD-1/PD-L1 Inhibitors or With Single-Agent Chemotherapy. Ferrara R, Mezquita L, Texier M, Lahmar J, Audigier-Valette C,Tessonnier L, Mazieres J, Zalcman G, Brosseau S, Le Moulec S,Leroy L, Duchemann B, Lefebvre C, Veillon R, Westeel V, Koscielny S, Champiat S, Ferté C, Planchard D, Remon J, Boucher ME,Gazzah A, Adam J, Bria E, Tortora G, Soria JC, Besse B, **Caramella C. 2018 Nov**[55]**.**

3/ Clarification de la définition d'hyperprogression

Contexte

Plusieurs équipes originaires de différents pays ont également tenté de montrer l'existence du phénomène d'hyperprogression (HPD) dans différents types tumoraux. Néanmoins, le phénomène étant nouveau et très controversé, une méta-analyse de ces différents patients était nécessaire.
Un des obstacles majeurs à la réalisation de celle-ci était la définition mathématique même de l'HPD. Chaque équipe a utilisé une méthode sensiblement identique (TGR et ΔTGR) mais différente dans la définition finale de ce qu'est l'HPD.

Nous avons donc mené une étude dans le but de clarifier l'influence de chacune des définitions différentes de l'HPD sur le chiffre d'incidence de l'HPD en appliquant les différentes définitions sur la population des 406 patients de notre étude princeps.

Matériel et méthode

Nous avons repris les analyses « RECIST » de la population de 406 patients traités par immunothérapie pour un cancer bronchique de l'étude précédentes et avons appliqué les 4 définitions d'HPD proposés par 4 autres équipes internationales (tableau 15).

	Champiat Nov 2016 Clin. Cancer Res. 2017	**Kato Mar 2017** Clin. Cancer Res. 2017	**Saâda-Bouzid Apr 2017** Ann. Oncol. 2017	**Singavi Sep 2017** Ann Oncol 2017	**Ferrara Nov 2018** JAMA Oncol 2018
Letter	A	B	C	D	E
Definition	RECIST progression & $TGR_{EXP}/TGR_{REF} > 2$	TTF < 2 months & RECIST > 50 % & Progression pace > 2	$TGK_{EXP}/TGK_{REF} > 2$	RECIST progression & RECIST > 50 % & $TGR_{EXP}/TGR_{REF} > 2$	RECIST progression & $TGR_{EXP} - TGR_{REF} > 50$
Reported HPD incidence	9 % (12/131)	4 % (6/155)	29 % (10/34)	5 patients	13.8 % (56/406)
Histological types	Various (Melanoma, 34%. Lung, 10%)	Various (Melanoma, 33%. NSCLC, 25%)	Head and Neck Squamous-Cell Carcinoma (HNSCC)	Various	NSCLC

Tableau 15: résumé des différentes définitions mathématiques d'HPD (A B C D E) dans 5 études différentes ainsi que les incidences calculées dans 5 cohortes différentes de patients. HPD = hyperprogressive disease, TGR = tumor growth rate, TTF = time to treatment failure, TGK = tumor growth kinetics, NSCLC = non small cell lung cancer[15,55–58].

Nous avons dans un premier temps modélisé de manière théorique quels « comportements tumoraux » étaient classifiés HPD avec les définitions différentes afin d'apprécier quels patients étaient exclus/inclus de la catégorie HPD en faisant varier la définition.

Nous avons ensuite calculé l'incidence de l'HPD dans la même population suivant chacune des définitions et utilisé l'index de similarité de Jaccard afin d'étudier la concordance de chaque définition.

Enfin, au vu des résultats préliminaires que nous avons obtenus, nous avons souhaité proposer une définition consensuelle d'HPD basée sur le pronostic des patients en associant la cinétique tumorale « accélérée » et la survie globale, ce qui avait le plus de sens clinique.

Résultats

La modélisation des définitions théoriques de chaque auteur est illustrée dans la figure 16.

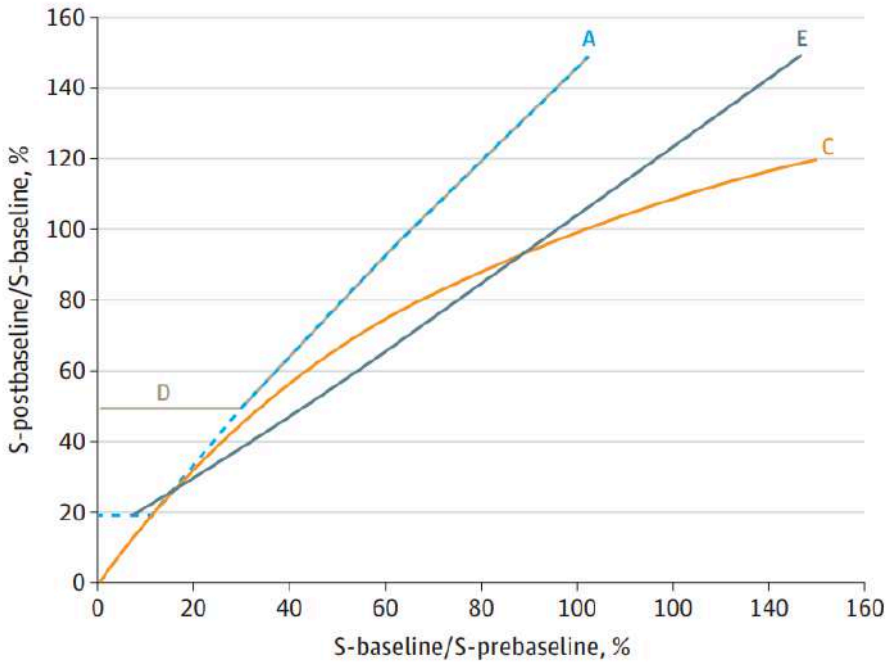


Figure 16 : Modélisation de la répartition des patients considérés HPD selon les différentes définitions. L'axe x correspond au TGR pré et l'axe y au TGRpost. Au-dessus de la ligne, les patients sont considérés HPD, au-dessous ils ne sont pas considérés HPD.

Lorsqu'on applique chacune des définitions à une population donnée, on obtient des incidences allant de 5.4 à 18.4% (22 à 75 patients) (tableau 16). L'application de l'index de similarité de Jaccard a montré qu'à l'échelon individuel, il ne s'agit pas systématiquement des mêmes patients puisque seuls 19 patients répondent à toutes les définitions (tableau 17).

	Définition A	Définition B	Définition C	Définition D	Définition E
Incidence	12.8% (52/406)	5.4% % (22/406)	18.4 % (75/406)	6.2 % (25/406)	13.8% (56/406)
Survie globale médiane HPD	5.1 months	3.4	6	5.1	4
Survie médiane globale non HPD	6.3 months	6.4	6.2	6.2	6.4
p-value	p = 0.45	p < 0.001	p = 0.62	p = 0.59	p = 0.14

Table 16 : incidence et survie globale des patients définis comme HPD et non HPD selon les différentes définitions.

	A (N=52)	B (N=22)	C (N=75)	D (N=25)	E (N=56)
A					
B	34.5 % (19)				
C	69.3 % (52)	27.6 % (21)			
D	48.1 % (25)	67.9 % (22)	33.3 % (25)		
E	68.8 % (44)	34.8 % (23)	59.8 % (49)	47.4 % (24)	

Tableau 17: index de similarité de Jaccard évaluant la concordance dans la constitution des groupes HPD/non HPD.

Enfin, afin de proposer une définition la plus objective possible qui ne repose pas que sur des chiffres de taille tumorale mais aussi sur la réalité clinique et la survie, nous avons corrélé la progression à la survie globale de chaque patient en partant du principe qu'une hyperprogression devenait cliniquement impactante si elle résultait en une survie plus courte des patients (figure 17). Une méthode d'évaluation de la valeur pronostique de chaque définition test a été réalisée et une analyse comparative de repère (comparative landmark analysis) a été faite.

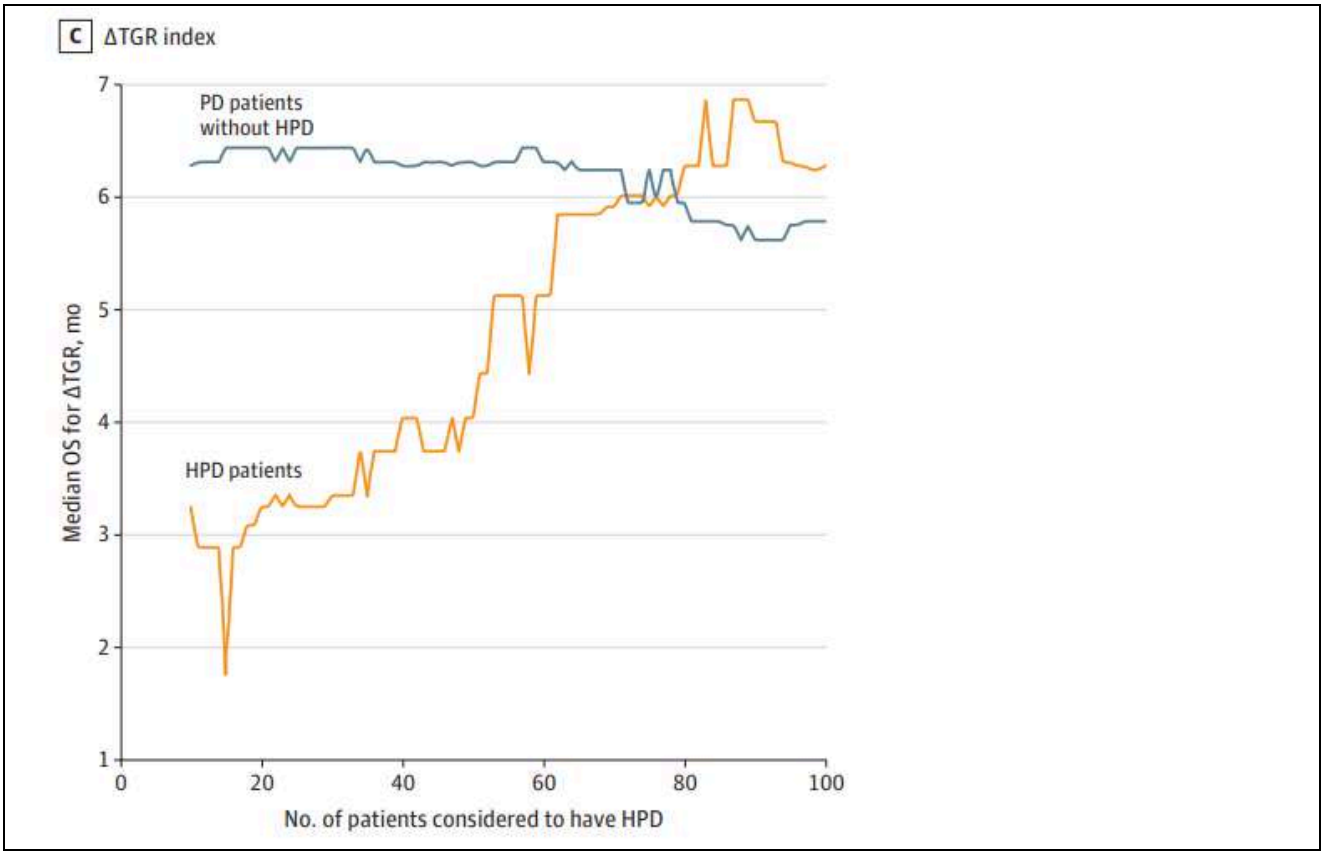


Figure 17 : Illustration de l'analyse comparative de repère réalisée pour tester les définitions d'HPD afin d'obtenir une séparation stable et en rapport avec les survies.

Ainsi nous avons montré sur cette cohorte particulière que la définition de HPD qui corrèle le mieux avec les pires survies et qui restait stable dans la modélisation était un ΔTGR>102.

En appliquant cette définition, l'incidence d'HPD était de 8.3%.

Publication dans JAMA Oncol . Clarification of Definitions of Hyperprogressive Disease During Immunotherapy for Non-Small Cell Lung Cancer. Kas B, Talbot H, Ferrara R, Richard C, Lamarque JP, Pitre-Champagnat S, Planchard D, Balleyguier C, Besse B, Mezquita L, Lassau N, **Caramella C**.. 2020 Jun 11;6(7):1039-46. doi: 10.1001/jamaoncol.2020.1634[59]

Perspectives

Les enjeux de l'évaluation tumorale deviennent de plus en plus importants, avec le développement de nouveaux traitements lourds et couteux, qui peuvent être responsables de modification du comportement tumoral parfois inattendus. Lorsque les indications de ces nouveaux traitements sont déterminées par des résultats statistiques de survie globale sur des populations randomisées entre nouveau traitement et ancien traitement « de référence », la place de l'imagerie est secondaire : si la population qui reçoit le nouveau traitement a une meilleure survie globale que la population témoin, alors le nouveau traitement sera validé dans cette indication (compte tenu des toxicités).

En revanche, l'imagerie devient un élément majeur pour l'identification de la progression (et donc de la poursuite ou non du traitement) et l'interprétation de nouveaux types de réponse comme la pseudo progression et l'hyperprogression. La question fondamentale, une fois les essais randomisés achevés, reste celle de la prédiction de la réponse. Quels patients vont réellement bénéficier du nouveau traitement et chez lesquels le traitement ne sera pas efficace, voire risque d'aggraver la situation ?

La réponse à cette question est complexe et reposera sur l'analyse de cohortes de plus en plus grandes et complexes avec l'intégration des données d'imagerie aux données cliniques et biologiques. Le développement de nouvelles collaborations avec des **équipes de recherche multidisciplinaires** permettra de définir des modèles cliniquement pertinents. Grâce aux techniques de deep learning, de nouvelles pistes s'ouvrent et des études sont actuellement en cours pour déterminer la place en clinique de la radiomique, en partenariat avec des starts-up ayant des compétences en intelligence artificielle.

Dans le chapitre suivant, j'ai abordé un autre axe de recherche concernant l'imagerie dans la prise en charge des patients atteints de cancers pulmonaires. Le sujet ne sera plus l'évaluation de la réponse à un traitement, mais la recherche de cible thérapeutique à l'aide de l'imagerie.

Chapitre 4 : L'imagerie comme biomarqueur prédictif de la présence d'une mutation oncogénique dans les cancers bronchiques métastatiques

Contexte

Depuis une quinzaine d'années, différentes anomalies génomiques au sein des cellules tumorales des adénocarcinomes bronchiques ont été identifiées. Les conséquences de ces découvertes ont permis de classifier de manière plus précise ce type de cancer et de proposer des traitements spécifiques ciblant ces anomalies. Le pronostic de patients porteurs de cancers métastatiques avec certaines anomalies génomiques s'est trouvé profondément modifié. A l'heure actuelle, la recherche des anomalies suivantes est recommandée chez tout patient diagnostiqué d'un adénocarcinome bronchique métastatique par l'ESMO: mutation EGFR, BRAF, réarrangement ALK, ROS1, NTRK[60]. La recherche d'autres anomalies est également recommandée par l'ASCO (mutation KRAS , NRAS, MET…).

La prévalence des différentes altérations est variable d'un pays à l'autre. Dans les pays caucasiens, les données de plusieurs organisations montrent une répartition relative illustrée dans la figure 18.

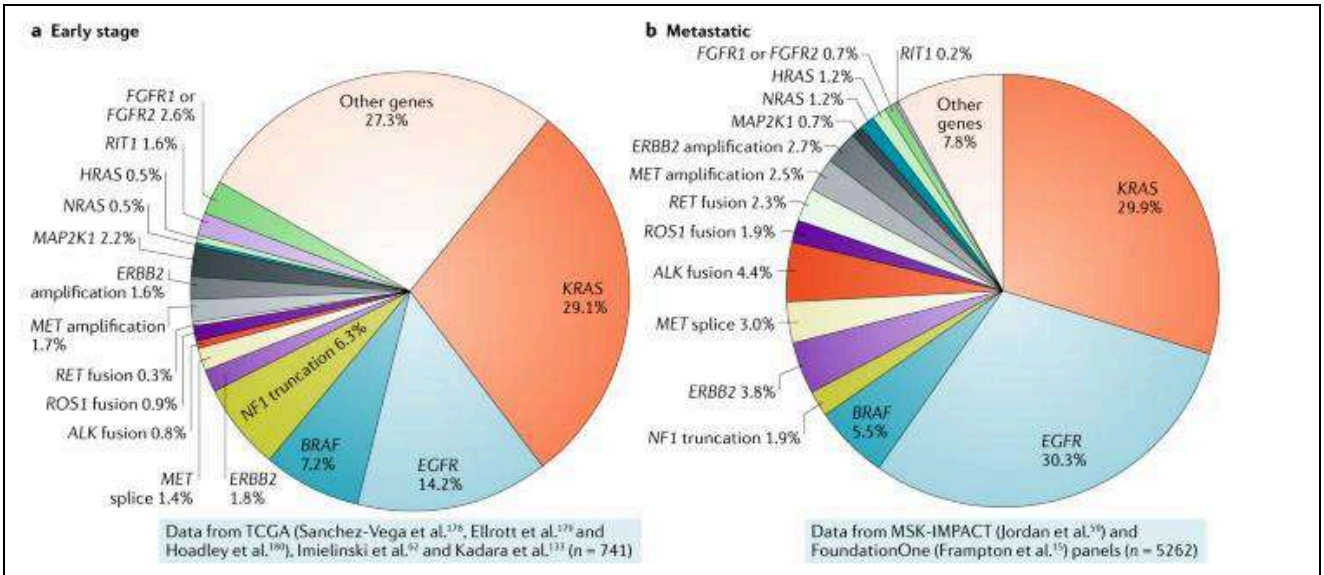


Figure 18 : Répartition des différentes altérations moléculaires (from ref [61]).

La recherche de ces anomalies nécessite du matériel tumoral, le plus souvent obtenu par des biopsies (sous fibroscopie ou par ponction trans-pariétale). Or environ 30% des patients ne pourront pas avoir de recherche d'anomalie génomique de leur tumeur en raison de matériel tumoral absent ou insuffisant, en rapport avec des difficultés à obtenir des prélèvements satisfaisants[62,63].

Certains phénotypes sont associés à certaines altérations génomiques : un cancer pulmonaire diagnostiqué chez une femme n'ayant jamais ou peu fumé suggère par exemple la présence d'une mutation du gène EGFR au sein des cellules tumorales.

En plus de ces arguments cliniques, nous avons plusieurs fois expérimenté que le tropisme métastatique de certaines tumeurs mutées semblait typique et différait d'une anomalie à l'autre. Nous avons cherché à savoir si le tropisme métastatique des adénocarcinomes pouvait révéler la présence d'une anomalie génomique d'intérêt chez nos patients.

Matériel et méthode

Nous avons mené une étude rétrospective chez les patients suivis à l'Institut Gustave Roussy pour un adénocarcinome bronchique métastatique entre janvier 2010 et mai 2017.

Les critères d'exclusion étaient l'absence de bilan d'imagerie complet (scanner thoraco-abdomino-pelvien, PET scanner au 18FDG et imagerie cérébrale scanner ou IRM), l'absence de recherche d'anomalie génomique, ou le diagnostic d'un deuxième cancer d'une autre origine.

Les critères cliniques classiques ont été relevés (âge, sexe, statut tabagique). Nous avons relu les imageries de manière exhaustive afin de déterminer quels organes présentaient des métastases (parmi os, cerveau, poumon, plèvre, surrénale, foie, péritoine, lymphangite pulmonaire, rate, parties molles muscle ou sous cutané, péricarde, rein, pancréas, thyroïde et adénopathies en dehors du médiastin et de la région sus claviculaire). Si plusieurs imageries étaient disponibles, nous avons analysé celle qui étaient la plus proche du premier diagnostic de métastase.

Les analyses moléculaires portaient sur la présence ou l'absence de détection d'anomalie dans les gènes EGFR, KRAS, BRAF et ALK. Les tumeurs qui étaient négatives pour la recherche de ces 4 anomalies étaient dénommées 4N (fourth negative).

Résultats

La figure 19 résume le flow-chart de l'inclusion ainsi que la composition des sous groupes EGFR, KRAS, BRAF, ALK et 4N.

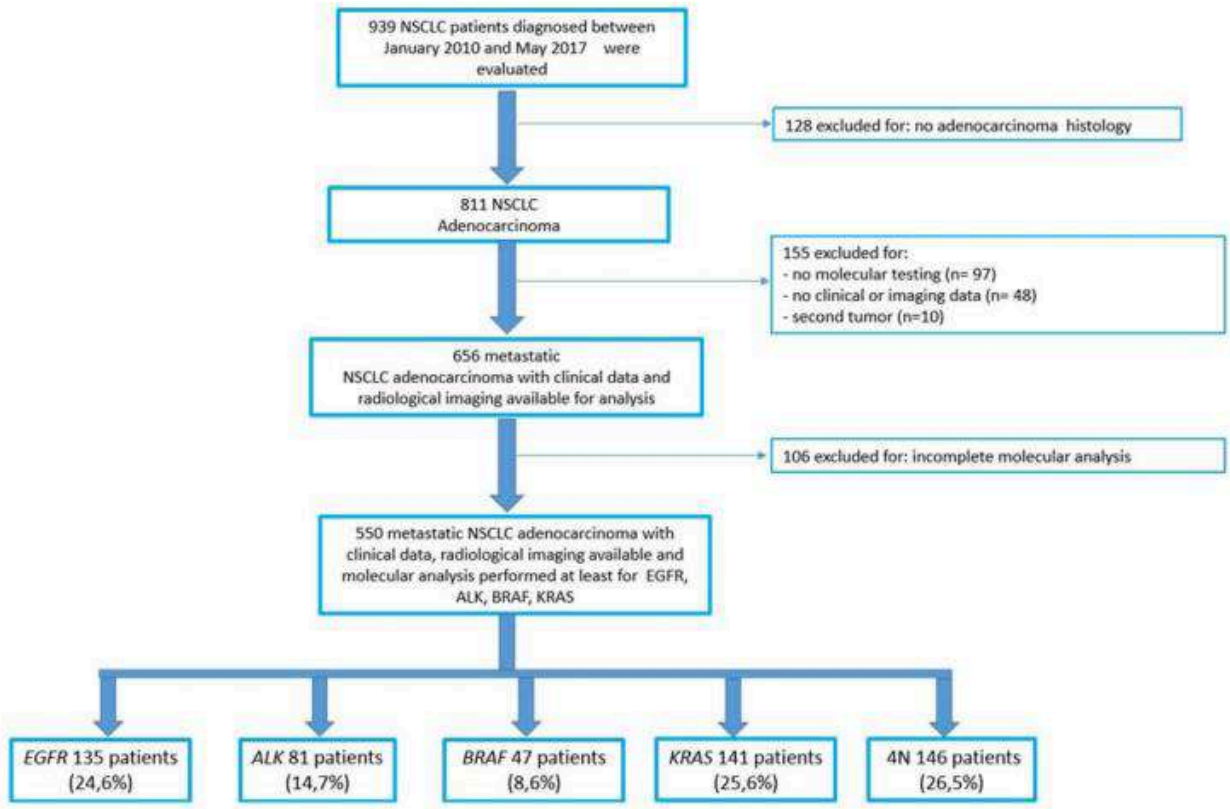


Figure 19 : Flow chart de l'étude et répartition des tumeurs par sous groupes moléculaires.

Au total, 550 patients ont été inclus (sex ratio 55/45%) avec une moyenne d'âge de 59 ans.

Le tableau 18 illustre les principales différences démographiques selon les sous-groupes. Les patientes non fumeuses étaient logiquement plus représentées dans le sous-groupe EGFR. Les patients du groupe BRAF étaient plus âgés, et ceux du groupe ALK plus jeunes.

	EGFR $n = 135$ (%)	*ALK* $n = 81$ (%)	*BRAF* $n = 47$ (%)	*KRAS* $n = 141$ (%)	*4N* $n = 146$ (%)
Gender					
Female	96 (71)***	39 (48)	21 (45)	52 (37)	60 (41)
Male	39 (29)	42 (52)	26 (55)	89 (63)	86 (59)
Median, range	59, 25–88	52***, 23–82	63*, 35–88	59, 30–83	59, 29–89
Smoking status					
Non-smoker	78 (58)***	37 (46)***	17 (36)*	6 (4)***	28 (19)
Former smoker	40 (30)	26 (32)	16 (34)	55 (39)	48 (33)
Current smoker	17 (13)	18 (22)	14 (30)	80 (57)	70 (48)

*p value compared with control group 4N < 0.05

***< 0.001

Tableau 18 : Caractéristiques démographiques des patients selon les différents sous groupes de tumeurs et leur anomalie moléculaire.

Concernant le tropisme métastatique, le tableau 19 en illustre la répartition.

Le site métastatique le plus fréquemment retrouvé chez les patients était l'os (44% des patients), le cerveau (33%) et le poumon (30%). Les métastases surrénaliennes, bien que très souvent citées comme « typiques » sont moins fréquentes (17% au total dans notre population).

	EGFR $n = 135$ (%)	*ALK* $n = 81$ (%)	*BRAF* $n = 47$ (%)	*KRAS* $n = 141$ (%)	*4N* $n = 146$ (%)
Bone	71 (53)	35 (43)	10 (21)*	66 (47)	61 (42)
Brain	48 (36)	34 (42)*	10 (21)	51 (36)	42 (29)
Lung	43 (32)	30 (37)*	14 (30)	41 (29)	35 (24)
Pleura	43 (32)*	19 (23%)	22 (47)***	23 (16)	29 (20)
Adrenal gland	8 (6)***	12 (15)	5 (11)	38 (27)	34 (23)
Lymph node	15 (11)*	17 (21)	7 (15)	20 (14)	33 (23)
Liver	26 (19)	15 (19)	6 (13)	24 (17)	20 (14)
Peritoneum	1 (1)	3 (4)	4 (9)	9 (6)	7 (5)
Lymphangitis	8 (6)*	6 (7)**	7 (15)***	3 (2)	1 (1)
Pericardium	2 (1)	3 (4)	5 (11)*	3 (2)	4 (3)
Soft tissue	2 (1)*	2 (2)	1 (2)	10 (7)	11 (8)

*p value compared with control group 4N < 0.05

**< 0.01

***< 0.001

Tableau 19 : Distribution respective des sites métastatiques en fonction de l'anomalie moléculaire identifiée dans la tumeur.

On constate qu'il existe un tropisme particulier pour les groupes EGFR, ALK et BRAF comparativement au groupe 4N. En revanche, le groupe KRAS n'est pas différent du groupe 4N.
Pour résumer, comparativement au groupe 4N (illustration figure 20):

- Les patients du groupe EGFR présentaient plus fréquemment des métastases pleurales et de la lymphangite carcinomateuse et moins souvent des métastases surrénaliennes, musculaires ou ganglionnaires.
- Les patients du groupe BRAF présentaient plus fréquemment une atteinte pleurale, péricardique et de la lymphangite et moins fréquemment des lésions osseuses
- Les patients du groupe ALK présentaient plus fréquemment des lésions cérébrales, pulmonaires et de la lymphangite.

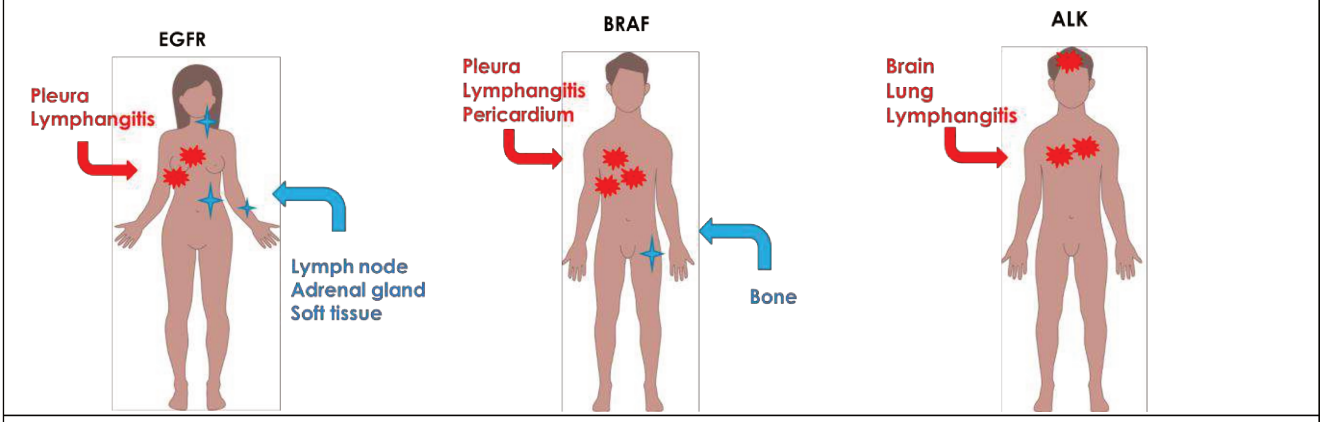


Figure 20 : illustration du tropisme métastatique préférentiel des patients avec une tumeur pulmonaire qui présente une anomalie du gène EGFR, BRAF et ALK.

Publication dans **European Radiology**. Dormieux A, Mezquita L, Cournede PH, Remon J, Tazdait M, Lacroix L, Rouleau E, Adam J, Bluthgen MV, Facchinetti F, Tselikas L, Aboubakar F, Naltet C, Lavaud P, Gazzah A, Le Pechoux C, Lassau N, Balleyguier C, Planchard D, Besse B, Caramella C. Association of metastatic pattern and molecular status in stage IV non-small cell lung cancer adenocarcinoma. 2020 Sep;30(9):5021-5028.64

Perspectives

Dans cette population sélectionnée de patients qui ont pu bénéficier d'une recherche d'anomalie moléculaire, on constate qu'il existe une relation génotype-phénotype sur le tropisme métastatique de chaque sous-groupe de cancer.

Cette découverte présente plusieurs intérêts :

- La mise en évidence d'un profil métastatique évoquant la présence d'une anomalie génomique avec des conséquences thérapeutiques chez un patient qui n'a pas été testé ou chez lequel il existe des difficultés à réaliser un prélèvement pourrait inciter plus fortement à réaliser ce prélèvement, ou à faire appel à des techniques innovantes et encore peu disponibles ou coûteuses (biopsies liquides par exemple).
- La connaissance de ces tropisme métastatiques pourraient améliorer l'analyse des examens d'imagerie standards afin d'améliorer la pertinence des comptes rendus, ou inciter à réaliser des explorations non encore standardisées (par exemple des IRM cérébrales systématiques chez les patients porteurs d'un adénocarcinome transloqué ALK).

A partir de cette cohorte de patients, une collaboration avec le laboratoire MICS de l'école Centrale Supelec et du Digital Lab (Paris) nous a permis de développer un algorithme de deep learning de « prédiction » : en renseignant l'âge, le sexe, le statut tabagique ainsi que les différents sites métastatiques retrouvés au bilan d'imagerie, on peut tester la probabilité de trouver une anomalie moléculaire.

Cet algorithme issu de l'analyse des 550 patients précédemment décrit, a pu être déployé au sein d'un site internet gratuit disponible à l'adresse **www.tactic-ct.fr** (illustration figure 21)**.** On peut ainsi renseigner les données cliniques et radiologiques d'un patient et obtenir un pourcentage de probabilité de présence d'une mutation d'interêt.

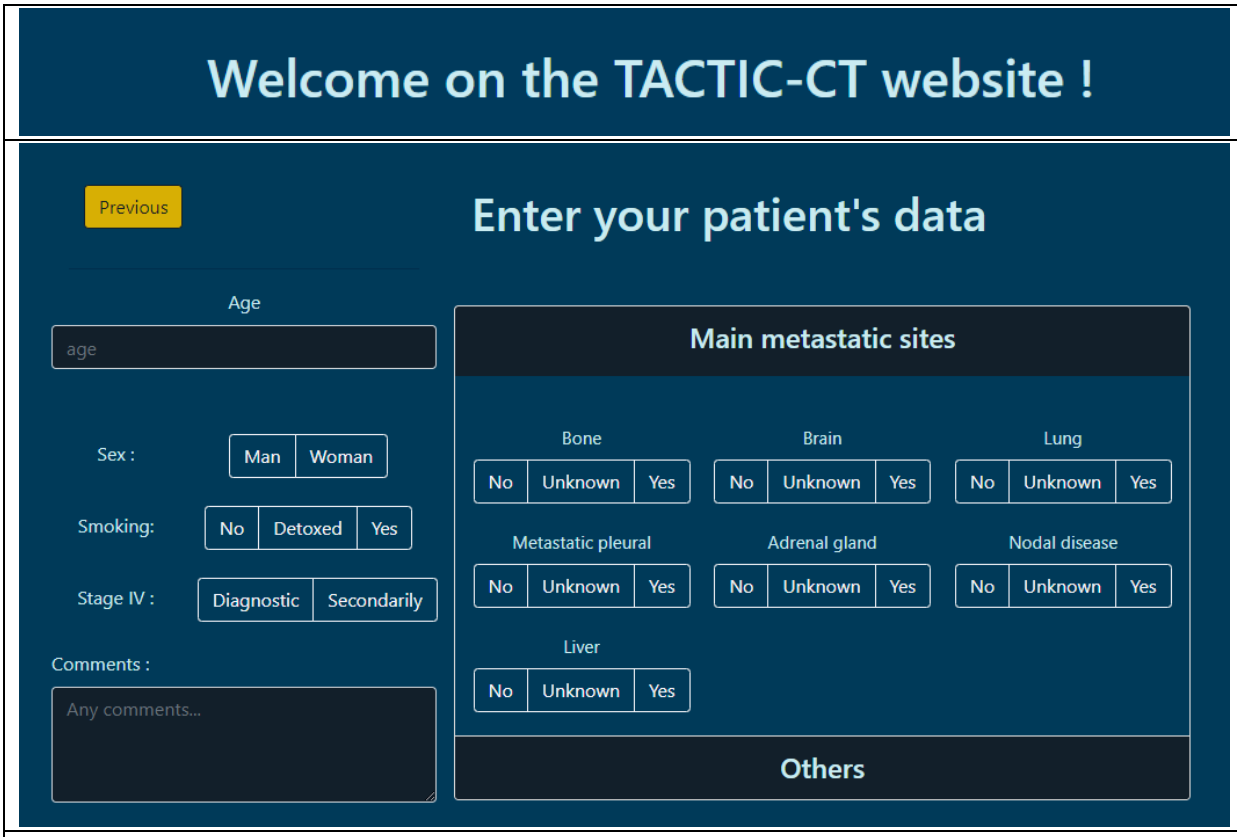


Figure 21 : capture d'écran du site internet tactic-ct.fr qui permet de tester la probabilité de trouver une anomalie moléculaire selon la présentation clinico-radiologique au diagnostic de stade métastatique.

Il a également été pensé comme un outil participatif, dans lequel il est possible de renseigner les données d'autres patients dont on connait les résultats d'analyse mutationnelle, afin d'alimenter l'algorithme qui pourra ensuite être entrainé à nouveau afin de fournir des résultats plus précis.

Conclusion - Synthèse

A l'issue de ce travail de thèse, il ressort que la place de l'imagerie comme biomarqueur en oncologie est un domaine extrêmement vaste et encore peu exploré, qui pourrait permettre d'améliorer considérablement la prise en charge des malades, tant sur la découverte d'anomalies ciblables que sur la prédiction de la réponse aux traitements ou encore la meilleure sélection de chaque patient pour un traitement particulier.

Ce travail permet également de poser un cadre majeur et préalable aux innovations souhaitées : l'importance de replacer l'imagerie dans un contexte méthodologique strict et ne pas considérer les images médicales comme des données magiques sans contrôle qualité. Ainsi, dans le chapitre 2, nous avons montré que la Radiomique devait dépasser ces limites méthodologiques avant de s'affirmer comme une vraie discipline.

La sélection d'indice de texture fiables est fondamentale et nous avons vu que, dans des conditions expérimentales simples sur fantôme, la répétabilité et la reproductibilité ne vont pas de soi, ni sur une même machine, ni sur plusieurs machines différentes. C'est aux radiologues d'informer les parties prenantes de ces enjeux et de faire pression sur les constructeurs afin qu'ils harmonisent une acquisition scannographique dédiée à la radiomique. Le consortium Cancer Core Europe œuvre dans ce sens, de même que d'autres équipes dans le monde.

Les études que j'ai menées sur la radiomique ne sont également que des études préliminaires, car l'objet qui a servi aux mesures est encore bien différent du corps humain, même si les inserts hétérogènes du fantôme avaient des densités plus proches des tissus humains. L'injection de produit de contraste indispensable pour l'obtention d'images anatomiques capable de visualiser les différents organes et d'éventuelles lésions métastatiques au sein des organes est utilisé en routine. Néanmoins, la répartition du produit de contraste dans le corps dépend de plusieurs facteurs tels que volume, débit et durée de l'injection, intervalle de temps entre le début de l'injection et l'acquisition des images, site de l'injection (voie veineuse périphérique/ centrale, côté). Ces éléments sont pris en compte en routine lorsque l'on veut mettre en évidence des anomalies vasculaires artérielles ou veineuses, ou des anomalies de rehaussement tardif, des voies urinaires, etc… car l'on sait que le produit sera présent dans tel ou tel compartiment. Quelles seront les conséquences sur l'analyse d'un volume d'intérêt en radiomique ?

Enfin, mes différents travaux de thèse soulignent l'importance de collaborer avec d'autres professionnels issus d'autres spécialités médicales, ingénieurs, mathématiciens, scientifiques, méthodologistes. L'importance d'évaluer correctement la réponse à un traitement n'est pas innée pour un radiologue et s'acquiert au contact des cliniciens qui doivent prendre des décisions lourdes en cas de progression radiologique. De même, l'observation de réponses atypiques, comme on a pu le

constater sous immunothérapie, a été la conséquence d'une expérience dans l'aspect des réponses typiques et les questionnements que nous avons eus sur l'aspect des courbes de survie dans les études chez des patients sous immunothérapie. Ce n'est que de cette façon que nous pourrons développer de nouvelles idées et trouver de nouvelles voies afin de faire avancer la recherche en imagerie médicale et plus globalement, la lutte contre le cancer.

Annexe 1 : Indices de texture, définitions mathématiques

❖ Histogramme des valeurs des intensités de pixel :

- Le Skewness de l'histogramme : mesure l'asymétrie de la distribution des valeurs d'intensité des voxels.

$$S = \frac{\frac{1}{E}\sum_i [H(i) - I_{mean}]^3}{\left[\sqrt{\frac{1}{E}\sum_i (H(i) - I_{mean})^2}\right]^3}$$

E : nombre de voxel dans le VOI (ou de pixel dans la ROI).

H(i) : nombre de voxels (ou de pixel) d'intensité i.

I_{mean} : intensité moyenne dans le VOI (ou ROI).

- Le Kurtosis de l'histogramme : mesure l'aplatissement de la distribution.

$$K = \frac{\frac{1}{E}\sum_i [H(i) - I_{mean}]^4}{\left[\frac{1}{E}\sum_i [H(i) - I_{mean}]^2\right]^2}$$

- L'*EntropyH* de la matrice de co-occurrence : mesure le désordre de la distribution de l'histogramme.

$$EntropyH = -\sum_i p(i).log_{10}(p(i))$$

- L'*EnergyH* de la matrice de co-occurrence : mesure l'uniformité de la distribution de l'histogramme.

$$EnergyH = \sum_i p(i)^2$$

❖ **La matrice de co-occurrence ou Gray-Level Co-occurrence Matrix (GLCM)**: permet de quantifier les relations entre paires de voxels pour différentes distances dans une direction donnée. Il en résulte une matrice carrée $G \times G$, avec G le nombre de niveaux de gris dans l'image après discrétisation. Chaque élément (i,j) de la matrice donne la probabilité d'avoir, dans la ROI considérée, deux voxels d'intensité i et j, séparés d'une distance *d* dans une direction donnée selon le vecteur *(d,θ)*, correspondant à un déplacement (Δx,Δy).

$$C_{\Delta x,\Delta y}(i,j) = \frac{1}{N \times M}\sum_i^{N-\Delta x}\sum_j^{M-\Delta y}\begin{cases}1\ si\ I(p,q) = i\ et\ I(p+\Delta x, q+\Delta y) = j\\ 0\ sinon\end{cases}$$

où *I*(*p, q*) représente le voxel de coordonnées (*p, q*) dans l'image *I* de taille *N*×*M*. Le plus souvent la distance *d* est choisie égale à 1. Le vecteur $\vec{d}$ permet de former toutes les combinaisons possibles entre

un voxel et ses 26 voisins en 3-dimensions. Exemple de matrice de co-occurrence :

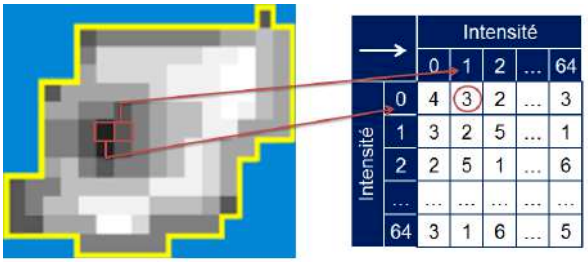


Figure 1: Illustration de la matrice de co-occurrence, issue de la thèse de Fanny Orlhac.

- Le *Contrast* de la matrice de co-occurrence: mesure le contraste dans l'image et augmente avec celui-ci.

$$Contrast = \sum_i \sum_j (i-j)^2 \times C(i,j)$$

$C(i,j)$ correspond à l'élément (i,j) de la matrice de co-occurrence.

- L'*Entropy* de la matrice de co-occurrence : mesure le désordre dans l'image et augmente lorsque la texture est aléatoire.

$$Entropy = -\sum_i \sum_j C(i,j) \times log_{10}(C(i,j))$$

- La *Dissimilarity* de la matrice de co-occurrence : mesure de la variation de gris dans l'image.

$$Dissimilarity = \sum_i \sum_j |i-j| \times C(i,j)$$

- L'*Homogeneity* de la matrice de co-occurrence : mesure l'homogénéité des paires de voxels de même niveau de gris.

$$Homogeneity = \sum_i \sum_j \frac{C(i,j)}{1+|i-j|}$$

- L'*Energy* de la matrice de co-occurrence : mesure l'uniformité des paires de voxels de même intesité.

$$\left(\sum_i \sum_j C(i,j)^2 \right)$$

- La *Correlation* de la matrice de co-occurrence : mesure la dépendance linéaire des niveaux de gris dans la matrice de co-occurrence.

$$Correlation = Average\ 13\ directions \sum_i \sum_j \frac{(i - \mu_i) \times (j - \mu_j) \times C(i,j)}{\sigma_i \times \sigma_j}$$

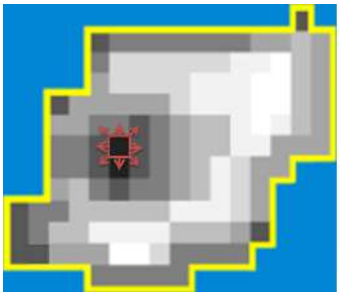
❖ **<u>La matrice de différences de niveaux de gris ou Neighborhood Grey-Level Difference Matrix (NGLDM)</u>** :

donne les différences de niveaux de gris entre un voxel et ses 26 voisins. La matrice résultante est de taille $G \times 2$.

D(i,1) ⇔ probabilité d'apparition du niveau i.

$$D(i,2) = \sum_p \sum_q \begin{cases} |\bar{M}(p,q) - i| \ si\ I(p,q) = i \\ 0\ sinon \end{cases}$$

où $\bar{M}(p,q)$ est la moyenne de l'intensité des 26 voisins du voxel de coordonnées (p,q). Exemple de matrice NGLDM :



		Prob.	Diff moy
Intensité	0	N(0)/N(Tot)	Σ |M(0) - 0|
	1	N(1)/N(Tot)	Σ |M(1) - 1|
	2	N(2)/N(Tot)	Σ |M(2) - 2|
	...	...	...
	64	N(64)/N(Tot)	Σ |M(64) - 64|

Figure 2: Illustration de la NGLDM, issue de la thèse de Fanny Orlhac.

- Le Contrast de la matrice de différences de niveaux de gris : mesure le contraste sur l'image.

$$Contrast = \left[\sum_i \sum_j D(i,1) \times D(j,1) \times (i-j)^2\right] \times \frac{\sum_i D(i,2)}{E \times G \times (G-1)}$$

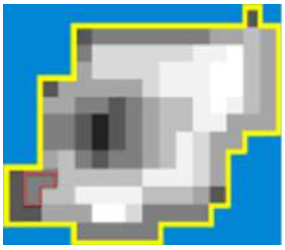
G : nombre de niveaux de gris.

❖ **<u>La matrice des longueurs des zones homogènes ou Grey-Level Zone Lenght Matrix (GLZLM) :</u>**

donne la taille des zones de voxel (ou pixel) de la même intensité et leur nombre dans le VOI (ou la ROI). La matrice GLZLM est de taille $G \times T$, avec G le nombre de niveaux de gris et T la taille de la plus grande zone homogène de voxels.

L'élément (i, j) correspond au nombre de zones de j voxels ayant une intensité i.

Exemple de matrice GLZLM :



Intensité	Taille des zones 1	2	3	...	36
0	14	5	2	...	4
1	8	6	18	...	2
2	5	3	11	...	1
...	...	...	...	...	...
64	1	2	9	...	0

Figure 3: Illustration de la GLZLM, issue de la thèse de Fanny Orlhac.

- Le Short Zone Emphasis (SZE) de la matrice des longueurs des zones homogènes ou Grey-Level Zone Lenght Matrix ou le Long Zone Emphasis (LZE) (GLZLM) : mesure la distribution des petites ou longues zones homogènes dans l'image.

$$SZE = \frac{1}{H}\sum_i\sum_j \frac{Z(i,j)}{j^2} \qquad LZE = \frac{1}{H}\sum_i\sum_j Z(i,j)\times j^2$$

H : nombre de zones homogènes dans le VOI.

$Z(i,j)$: élément (i,j) de la matrice (GLZLM).

- Le Low Grey-level Zone Emphasis (LGZE) ou le High Grey-level Zone Emphasis (HGZE) de la GLZLM: mesure de la distribution des zones de bas ou haut niveau de gris.

$$LGZE = \frac{1}{H}\sum_i\sum_j \frac{Z(i,j)}{i^2} \qquad HGZE = \frac{1}{H}\sum_i\sum_j Z(i,j)\times i^2$$

- Le Short-Zone Low Grey-level Emphasis (SZLGE) ou le Short-Zone High Grey-level Emphasis (SZHGE) de la GLZLM : mesure de la distribution des courtes zones homogènes de bas ou haut niveau de gris.

$$SZLGE = \frac{1}{H}\sum_i\sum_j \frac{Z(i,j)}{i^2\times j^2} \qquad SZHGE = \frac{1}{H}\sum_i\sum_j \frac{Z(i,j)\times i^2}{j^2}$$

- Le Long-Zone Low Grey-level Emphasis (LZLGE) ou le Long-Zone High Grey-level Emphasis (LZHGE) de la GLZLM : mesure la distribution des longues zones homogènes de bas ou haut niveau de gris.

$$LZLGE = \frac{1}{H}\sum_i \sum_j \frac{Z(i,j) \times i^2}{j^2} \qquad LZHGE = \frac{1}{H}\sum_i \sum_j Z(i,j) \times i^2 \times j^2$$

- Le Grey-Level Non-Uniformity for zone (GLNUz) ou le Zone Lenght Non-Uniformity de la GLZLM : mesure la non uniformité des niveaux de gris ou de la longueur des zones homogènes.

$$GLNUz = \frac{1}{H}\sum_i \left(\sum_j Z(i,j)\right)^2 \qquad ZLNU = \frac{1}{H}\sum_j \left(\sum_i Z(i,j)\right)^2$$

- Le Zone Percentage (ZP) de la GLZLM : mesure l'homogénéité des zones homogènes.

$$ZP = \frac{H}{\sum_i \sum_j (j \times Z(i,j))}$$

❖ **<u>La matrice des longueurs des séries homogènes ou Grey-Level Run Lenght Matrix (GLRLM) :</u>**

Même indices que dans la GLZLM mais en remplaçant le Z(i,j) en R(i,j) et H devient le nombre de séries homogènes dans le VOI.

Concernant les indices de texture calculés dans le logiciel Lifex, une page web spécifique est consultable à l'adresse suivante :
https://lifexsoft.org/index.php/resources/19-texture/radiomic-features

BIBLIOGRAPHIE

1. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009;45(2):228-247. doi:10.1016/j.ejca.2008.10.026

2. Hanahan D, Weinberg RA. Hallmarks of Cancer: The Next Generation. *Cell*. 2011;144(5):646-674. doi:10.1016/j.cell.2011.02.013

3. Thiam R, Fournier LS, Trinquart L, et al. Optimizing the size variation threshold for the CT evaluation of response in metastatic renal cell carcinoma treated with sunitinib. *Ann Oncol*. 2010;21(5):936-941. doi:10.1093/annonc/mdp466

4. Eggermont A, Robert C, Soria JC, Zitvogel L. Harnessing the immune system to provide long-term survival in patients with melanoma and other solid tumors. *OncoImmunology*. 2014;3(1):e27560. doi:10.4161/onci.27560

5. Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther*. 2001;69(3):89-95. doi:10.1067/mcp.2001.113989

6. Eslami-S Z, Cortés-Hernández LE, Alix-Panabières C. The Metastatic Cascade as the Basis for Liquid Biopsy Development. *Front Oncol*. 2020;10:1055. doi:10.3389/fonc.2020.01055

7. deSouza NM, Winfield JM, Waterton JC, et al. Implementing diffusion-weighted MRI for body imaging in prospective multicentre trials: current considerations and future perspectives. *Eur Radiol*. 2018;28(3):1118-1131. doi:10.1007/s00330-017-4972-z

8. Sorace AG, Elkassem AA, Galgano SJ, et al. Imaging for Response Assessment in Cancer Clinical Trials. *Semin Nucl Med*. 2020;50(6):488-504. doi:10.1053/j.semnuclmed.2020.05.001

9. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-577. doi:10.1148/radiol.2015151169

10. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer Oxf Engl 1990*. 2012;48(4):441-446. doi:10.1016/j.ejca.2011.11.036

11. Fornacon-Wood I, Mistry H, Ackermann CJ, et al. Reliability and prognostic value of radiomic features are highly dependent on choice of feature extraction platform. *Eur Radiol*. 2020;30(11):6241-6250. doi:10.1007/s00330-020-06957-9

12. Eggermont AMM, Robert C. New drugs in melanoma: it's a whole new world. *Eur J Cancer Oxf Engl 1990*. 2011;47(14):2150-2157. doi:10.1016/j.ejca.2011.06.052

13. Forde PM, Kelly RJ, Brahmer JR. New Strategies in Lung Cancer: Translating Immunotherapy into Clinical Practice. *Clin Cancer Res*. 2014;20(5):1067-1073. doi:10.1158/1078-0432.CCR-13-0731

14. Wolchok JD, Hoos A, O'Day S, et al. Guidelines for the Evaluation of Immune Therapy Activity in Solid Tumors: Immune-Related Response Criteria. *Clin Cancer Res*. 2009;15(23):7412-7420. doi:10.1158/1078-0432.CCR-09-1624

15. Champiat S, Dercle L, Ammari S, et al. Hyperprogressive Disease Is a New Pattern of Progression in Cancer Patients Treated by Anti-PD-1/PD-L1. *Clin Cancer Res*. 2017;23(8):1920-1928. doi:10.1158/1078-0432.CCR-16-1741

16. O'Connor JPB, Aboagye EO, Adams JE, et al. Imaging biomarker roadmap for cancer studies. *Nat Rev Clin Oncol*. 2017;14(3):169-186. doi:10.1038/nrclinonc.2016.162

17. Fave X, Cook M, Frederick A, et al. Preliminary investigation into sources of uncertainty in quantitative imaging features. *Comput Med Imaging Graph*. 2015;44:54-61. doi:10.1016/j.compmedimag.2015.04.006

18. Jirák D, Dezortová M, Hájek M. Phantoms for texture analysis of MR images. Long-term and multi-center study. *Med Phys*. 2004;31(3):616-622. doi:10.1118/1.1646231

19. Nioche C, Orlhac F, Boughdad S, et al. LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res*. 2018;78(16):4786-4789. doi:10.1158/0008-5472.CAN-18-0125

20. Fave X, Mackin D, Yang J, et al. Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? *Med Phys*. 2015;42(12):6784-6797. doi:10.1118/1.4934826

21. Hunter LA, Krafft S, Stingo F, et al. High quality machine-robust image features: identification in nonsmall cell lung cancer computed tomography images. *Med Phys*. 2013;40(12):121916. doi:10.1118/1.4829514

22. Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. *Transl Oncol*. 2014;7(1):72-87. doi:10.1593/tlo.13844

23. Caramella C, Allorant A, Orlhac F, et al. Can we trust the calculation of texture indices of CT images? A phantom study. *Med Phys*. 2018;45(4):1529-1536. doi:10.1002/mp.12809

24. Al-Kadi OS. Assessment of texture measures susceptibility to noise in conventional and contrast enhanced computed tomography lung tumour images. *Comput Med Imaging Graph Off J Comput Med Imaging Soc*. 2010;34(6):494-503. doi:10.1016/j.compmedimag.2009.12.011

25. Berenguer R, Pastor-Juan MDR, Canales-Vázquez J, et al. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology*. 2018;288(2):407-415. doi:10.1148/radiol.2018172361

26. Buch K, Li B, Qureshi MM, Kuno H, Anderson SW, Sakai O. Quantitative Assessment of Variation in CT Parameters on Texture Features: Pilot Study Using a Nonanatomic Phantom. *AJNR Am J Neuroradiol*. 2017;38(5):981-985. doi:10.3174/ajnr.A5139

27. He L, Huang Y, Ma Z, Liang C, Liang C, Liu Z. Effects of contrast-enhancement, reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci Rep*. 2016;6:34921. doi:10.1038/srep34921

28. Kim H, Park CM, Lee M, et al. Impact of Reconstruction Algorithms on CT Radiomic Features of Pulmonary Tumors: Analysis of Intra- and Inter-Reader Variability and Inter-Reconstruction Algorithm Variability. *PloS One*. 2016;11(10):e0164924. doi:10.1371/journal.pone.0164924

29. Larue RTHM, van Timmeren JE, de Jong EEC, et al. Influence of gray level discretization on radiomic feature stability for different CT scanners, tube currents and slice thicknesses: a comprehensive phantom study. *Acta Oncol Stockh Swed*. 2017;56(11):1544-1553. doi:10.1080/0284186X.2017.1351624

30. Lu L, Ehmke RC, Schwartz LH, Zhao B. Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings. *PloS One*. 2016;11(12):e0166550. doi:10.1371/journal.pone.0166550

31. Mackin D, Fave X, Zhang L, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Invest Radiol*. 2015;50(11):757-765. doi:10.1097/RLI.0000000000000180

32. Mackin D, Fave X, Zhang L, et al. Harmonizing the pixel size in retrospective computed tomography radiomics studies. *PloS One*. 2017;12(9):e0178524. doi:10.1371/journal.pone.0178524

33. Mackin D, Ger R, Dodge C, et al. Effect of tube current on computed tomography radiomic features. *Sci Rep*. 2018;8(1):2354. doi:10.1038/s41598-018-20713-6

34. Mahmood U, Apte AP, Deasy JO, Schmidtlein CR, Shukla-Dave A. Investigating the Robustness Neighborhood Gray Tone Difference Matrix and Gray Level Co-occurrence Matrix Radiomic Features on Clinical Computed Tomography Systems Using Anthropomorphic Phantoms: Evidence From a Multivendor Study. *J Comput Assist Tomogr*. 2017;41(6):995-1001. doi:10.1097/RCT.0000000000000632

35. Midya A, Chakraborty J, Gönen M, Do RKG, Simpson AL. Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility. *J Med Imaging Bellingham Wash*. 2018;5(1):011020. doi:10.1117/1.JMI.5.1.011020

36. Shafiq-Ul-Hassan M, Zhang GG, Latifi K, et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med Phys*. 2017;44(3):1050-1062. doi:10.1002/mp.12123

37. Solomon J, Mileto A, Nelson RC, Roy Choudhury K, Samei E. Quantitative Features of Liver Lesions, Lung Nodules, and Renal Stones at Multi-Detector Row CT Examinations: Dependency on Radiation Dose and Reconstruction Algorithm. *Radiology*. 2016;279(1):185-194. doi:10.1148/radiol.2015150892

38. Yang J, Zhang L, Fave XJ, et al. Uncertainty analysis of quantitative imaging features extracted from contrast-enhanced CT in lung tumors. *Comput Med Imaging Graph Off J Comput Med Imaging Soc*. 2016;48:1-8. doi:10.1016/j.compmedimag.2015.12.001

39. Zhao B, Tan Y, Tsai WY, Schwartz LH, Lu L. Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study. *Transl Oncol*. 2014;7(1):88-93. doi:10.1593/tlo.13865

40. Zwanenburg A, Vallières M, Abdalah MA, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology*. 2020;295(2):328-338. doi:10.1148/radiol.2020191145

41. Espinasse M, Pitre-Champagnat S, Charmettant B, et al. CT Texture Analysis Challenges: Influence of Acquisition and Reconstruction Parameters: A Comprehensive Review. *Diagnostics*. 2020;10(5):258. doi:10.3390/diagnostics10050258

42. Eggermont AMM, Apolone G, Baumann M, et al. Cancer Core Europe: A translational research infrastructure for a European mission on cancer. *Mol Oncol*. 2019;13(3):521-527. doi:10.1002/1878-0261.12447

43. Calvo F, Apolone G, Baumann M, et al. Cancer Core Europe: A European cancer research alliance realizing a research infrastructure with critical mass and programmatic approach to cure cancer in the 21st century. *Eur J Cancer*. 2018;103:155-159. doi:10.1016/j.ejca.2018.08.023

44. A.I. Versus M.D. | The New Yorker. Accessed May 10, 2021. https://www.newyorker.com/magazine/2017/04/03/ai-versus-md

45. van Leeuwen KG, Schalekamp S, Rutten MJCM, van Ginneken B, de Rooij M. Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *Eur Radiol*. Published online April 15, 2021. doi:10.1007/s00330-021-07892-z

46. Hosny A, Parmar C, Quackenbush J, Schwartz LH, Aerts HJWL. Artificial intelligence in radiology. *Nat Rev Cancer*. 2018;18(8):500-510. doi:10.1038/s41568-018-0016-5

47. European Society of Radiology (ESR). What the radiologist should know about artificial intelligence - an ESR white paper. *Insights Imaging*. 2019;10(1):44. doi:10.1186/s13244-019-0738-2

48. Mellman I, Coukos G, Dranoff G. Cancer immunotherapy comes of age. *Nature*. 2011;480(7378):480-489. doi:10.1038/nature10673

49. Di Giacomo AM, Danielli R, Guidoboni M, et al. Therapeutic efficacy of ipilimumab, an anti-CTLA-4 monoclonal antibody, in patients with metastatic melanoma unresponsive to prior systemic treatments: clinical and immunological evidence from three patient cases. *Cancer Immunol Immunother CII*. 2009;58(8):1297-1306. doi:10.1007/s00262-008-0642-y

50. Seymour L, Bogaerts J, Perrone A, et al. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol*. 2017;18(3):e143-e152. doi:10.1016/S1470-2045(17)30074-8

51. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. *N Engl J Med*. 2015;373(17):1627-1639. doi:10.1056/NEJMoa1507643

52. Gomez-Roca C, Koscielny S, Ribrag V, et al. Tumour growth rates and RECIST criteria in early drug development. *Eur J Cancer Oxf Engl 1990*. 2011;47(17):2512-2516. doi:10.1016/j.ejca.2011.06.012

53. Champiat S, Ferrara R, Massard C, et al. Hyperprogressive disease: recognizing a novel pattern to improve patient management. *Nat Rev Clin Oncol*. 2018;15(12):748-762. doi:10.1038/s41571-018-0111-2

54. Ferté C, Fernandez M, Hollebecque A, et al. Tumor growth rate is an early indicator of antitumor drug activity in phase I clinical trials. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2014;20(1):246-252. doi:10.1158/1078-0432.CCR-13-2098

55. Ferrara R, Mezquita L, Texier M, et al. Hyperprogressive Disease in Patients With Advanced Non-Small Cell Lung Cancer Treated With PD-1/PD-L1 Inhibitors or With Single-Agent Chemotherapy. *JAMA Oncol*. 2018;4(11):1543-1552. doi:10.1001/jamaoncol.2018.3676

56. Kato S, Goodman A, Walavalkar V, Barkauskas DA, Sharabi A, Kurzrock R. Hyperprogressors after Immunotherapy: Analysis of Genomic Alterations Associated with Accelerated Growth Rate. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2017;23(15):4242-4250. doi:10.1158/1078-0432.CCR-16-3133

57. Saâda-Bouzid E, Defaucheux C, Karabajakian A, et al. Hyperprogression during anti-PD-1/PD-L1 therapy in patients with recurrent and/or metastatic head and neck squamous cell carcinoma. *Ann Oncol Off J Eur Soc Med Oncol*. 2017;28(7):1605-1611. doi:10.1093/annonc/mdx178

58. Singavi AK, Menon S, Kilari D, et al. Predictive biomarkers for hyper-progression (HP) in response to immune checkpoint inhibitors (ICI) – analysis of somatic alterations (SAs). *Ann Oncol*. 2017;28:v405. doi:10.1093/annonc/mdx376.006

59. Kas B, Talbot H, Ferrara R, et al. Clarification of Definitions of Hyperprogressive Disease During Immunotherapy for Non–Small Cell Lung Cancer. *JAMA Oncol*. 2020;6(7):1039. doi:10.1001/jamaoncol.2020.1634

60. Planchard D, Popat S, Kerr K, et al. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2018;29:iv192-iv237. doi:10.1093/annonc/mdy275

61. Skoulidis F, Heymach JV. Co-occurring genomic alterations in non-small-cell lung cancer biology and therapy. *Nat Rev Cancer*. 2019;19(9):495-509. doi:10.1038/s41568-019-0179-8

62. Kembou Nzale S, Weeks WB, Ouafik L, et al. Inequity in access to personalized medicine in France: Evidences from analysis of geo variations in the access to molecular profiling among advanced non-small-cell lung cancer patients: Results from the IFCT Biomarkers France Study. *PloS One*. 2020;15(7):e0234387. doi:10.1371/journal.pone.0234387

63. Zugazagoitia J, Ramos I, Trigo JM, et al. Clinical utility of plasma-based digital next-generation sequencing in patients with advance-stage lung adenocarcinomas with insufficient tumor samples for tissue genotyping. *Ann Oncol Off J Eur Soc Med Oncol*. 2019;30(2):290-296. doi:10.1093/annonc/mdy512

64. Dormieux A, Mezquita L, Cournede PH, et al. Association of metastatic pattern and molecular status in stage IV non-small cell lung cancer adenocarcinoma. *Eur Radiol*. 2020;30(9):5021-5028. doi:10.1007/s00330-020-06784-y

Article Type: Research Article

Can we trust the calculation of texture indices of CT images? A phantom study.

Caramella C[1,2] MD, Allorant A[3] MSc, Orlhac F[4] PhD, Bidault F[1,2]MD, Asselain B[2,5] PhD, Ammari S[1] MD, Jaranowski P[1], Moussier A[1] MSc, Balleyguier C[1,2] MD PhD, Lassau N[2,5]Pr, Pitre-Champagnat S[2] PhD.

1 Department of Imaging. Gustave Roussy Cancer Campus. 114 rue Edouard Vaillant. 94805 Villejuif cedex, France.

2 IR4M-UMR8081, CNRS, Univ Paris Sud, University Paris Saclay, Rue Ampère. 91405 Orsay cedex, France.

3 Department of Biostatistics. Gustave Roussy Cancer Campus. 114 rue Edouard Vaillant. 94805 Villejuif, France.

4 Imagerie Moléculaire In Vivo, INSERM, CEA, CNRS, University Paris Sud, University Paris Saclay, CEA, Orsay, France. INSERM, U1030, Villejuif, France

5 Research Department. Gustave Roussy Cancer Campus. 114 avenue Edouard Vaillant. 94805 Villejuif, France

Corresponding author

Caramella Caroline. Department of imaging. Gustave Roussy Cancer Campus. 114 rue Edouard Vaillant. 94805 Villejuif cedex. Phone number +33 42116321. Fax number +3342115495

Email address: caroline.caramella@gustaverousy.fr

Running head: CT texture analysis as an imaging biomarker

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/mp.12809

Abstract

Purpose: Texture analysis is an emerging tool in the field of medical imaging analysis. However, many issues have been raised in terms of its use in assessing patient images and it is crucial to harmonize and standardize this new imaging measurement tool. This study was designed to evaluate the reliability of texture indices of CT images on a phantom including a reproducibility study, to assess the discriminatory capacity of indices potentially relevant in CT medical images and to determine their redundancy.

Methods: For the reproducibility and discriminatory analysis, eight identical CT acquisitions were performed on a phantom including one homogeneous insert and two close heterogeneous inserts. Texture indices were selected for their high reproducibility and capability of discriminating different textures. For the redundancy analysis, 39 acquisitions of the same phantom were performed using varying acquisition parameters and a correlation matrix was used to explore the 2 x 2 relationships. LIFEx software was used to explore 34 different parameters including first order and texture indices.

Results: Only eight indices out of 34 exhibited high reproducibility and discriminated textures from each other. Skewness and kurtosis from histogram were independent from the six other indices but were inter-correlated, the other six indices correlated in diverse degrees (entropy, dissimilarity and contrast of the co-occurrence matrix, contrast of the Neighborhood Grey Level difference matrix, SZE, ZLNU of the Grey-Level Size Zone Matrix)

Conclusions: Care should be taken when using texture analysis as a tool to characterize CT images because changes in quantification may be primarily due to internal variability rather than from real physio-pathological effects. Some textural indices appear to be sufficiently reliable and capable to discriminate close textures on CT images.

Key words: texture analysis; computed tomography; phantom study,

INTRODUCTION

Texture is defined as the quantification of the spatial distribution of repeating patterns. It is used to translate the homogeneous or non-homogenous appearance of the surface of an object on an image. It is also defined as a two-dimensional phenomenon: the first dimension is a description of the basic elements, or the "primitive" (the pattern), from which the texture is formed; the second dimension relates to the description of the spatial organization of these primitives.[1]

Radiomics is a relatively new discipline for the clinical integration of the quantitative features of digital medical images as determined through mathematical analysis. One potential quantitative image feature of clinical significance is tumor tissue heterogeneity. Many studies have already explored the contribution of texture analysis in the field of oncology, addressing a range of topics (diagnosis, prognosis, correlation with histological or biological characteristics of the tumors) with wide-ranging results[2,3]. The main difficulty associated with this area is that there are currently as many ways of calculating texture indices as there are research groups focusing on texture analysis.

Several issues have already been raised: the definition of texture algorithms, which texture indices are relevant in medical images and the lack of reproducibility of texture calculations[4-7]. It is crucial to the success of this methodology to harmonize the procedures and textural features and to identify any inadequate parameters which could jeopardize the chances of obtaining useful results in clinical practice.

Our study is a first step towards the standardization of this new imaging measurement tool. It was designed to evaluate the reliability of texture indices of CT images on a phantom including a reproducibility study, to assess the discriminatory capacity of indices potentially relevant in CT medical images, and to determine their redundancy.

MATERIALS AND METHODS

Phantom

We built a phantom based on the commercially available CIRS® Electron Density Phantom Model 062M (Norfolk, Virginia, USA) which allows adequate maneuverability and the option of filling 17 holes with commercial homogeneous plugs as well as with “home-made” inserts.

The cylindrical phantom dimensions are 180 mm in diameter and 50 mm in thickness. Plug dimension is 30 mm in diameter and 50 mm in thickness. A homogenous plug with the same density as that of muscle (physical density of 1.06 g/cc) was used (named C1). In order to mimic human tissue, we designed two additional heterogeneous mixtures based on Ecoflex® (BASF, Ludwigshafen, Germany) which is a stable polymer, composed of two pure carbon fragments in different proportions randomly arranged in an agarose support (named C2 and C3), these two inserts have the advantage of having a similar range of physical density and histograms as those found in human soft tissue, such as tumoral lung tissues. Figure 1 shows the inserts and the corresponding image of a CT acquisition slice.

Reproducibility study

Eight identical CT acquisitions were made with a Discovery CT750 HD with 64 multi-detector arrays (GE Healthcare, Milwaukee, WI) using parameters routinely used for standard thoracic CT acquisition in patients treated in our center: voltage 120 kVp; x-ray tube current 150 mAs; 1.25 mm slice thickness; 1.375 pitch; rotation time 0.6 s, field of view 360 mm; reconstruction algorithm ASIR 0%; standard filter. Images were acquired sequentially and the phantom set-up was not altered.

Correlation of the texture index

For assessment of redundant indices, we performed 39 supplementary CT acquisitions on the same phantom and the same CT device, which differed for the acquisition parameters (voltage, intensity, slice thickness, pitch, rotation time, field of view, reconstruction algorithm, and filter (Table 1).

Texture analysis

Index extraction was performed with LIFEx (*Local Image Features Extraction*) open access software (http://www.lifexsoft.org; Orsay, France)[8] which calculates the minimum and maximum density values, as well as the mean and standard deviation of the grey values of the region of interest (ROI). From these primary calculations, the software accesses histogram values (skewness, kurtosis, entropyH, energyH), the co-occurrence matrix values (homogeneity, energy, contrast, correlation, entropy, dissimilarity), the Grey-Level Run Length Matrix GLRLM (SRE/LRE, LGRE / HGRE, SRLGE/ SRHGE, LRLGE/LRHGE, GLNU/RLNU, RP), the Neighborhood Grey Level difference matrix NGLDM (coarseness,

busyness, contrast) and the Grey-Level Size Zone Matrix GLSZM (SZE / LZE, LGZE / HGZE, SZLGE / SZHGE, LZLGE/ LZHGE, GLNUz / ZLNU, ZP).

The description of texture indices is provided in Supplemental Digital Content 1. An additional absolute discretization pixel intensity in Hounsfield units (HU) was applied to promote differentiation of intensities before texture index extraction[9]. We chose 200 gray values, by steps of 10 HU, between the extreme values of -1000 and +1000 HU. Two-dimensional circular ROI were drawn on each of the three inserts. The size of these ROI was identical for each insert (458 mm², 758 pixels) (figure 2).

Statistical analyses

The eight CT acquisitions resulted in eight tables of 34 texture indices from the three ROI. The reproducibility of the texture analysis was estimated from the coefficient of variation (CV). CVs were calculated for each texture index according to the following formula:

$$cv = \frac{\sigma}{\mu}$$

where σ is the standard deviation and μ the mean value of the data sample.

Variation in CV relating to imaging biomarkers is not standard[10], in this study, a CV < 20% was considered acceptable.

Dot plots were performed to show the dispersion of values of the different indices among C1, C2 and C3. A non-parametric Mann-Whitney test was used to exclude the indices which could not differentiate C1 from C2/C3, and C2 from C3. Indices that would be unable to distinguish a homogeneous area from a heterogeneous area were considered to be of no interest in the assessment of tumoral heterogeneity, as were indices that would not be able to distinguish two “visible to the naked eye” heterogeneous areas.

A Pearson correlation between the texture indices was assessed to evaluate their redundancy on the 39 different CT acquisitions. A correlation matrix was used to explore the 2 x 2 relationships between the retained parameters. Principal component analysis was used to represent the different retained indices and their pattern of correlation on a simple graph. Principal component analysis is a multivariate technique that analyzes a data table in which observations are described by several inter-correlated quantitative dependent variables. Its goal is to extract the important information from the table, to represent it as a set of new orthogonal variables called principal components, and to display the pattern of similarity of the observations and of the variables as points in maps[9].

RESULTS

The characteristics of intensities (in HU) obtained within each ROI (C1, C2 and C3) are reported in Table 2. As expected, the homogeneous ROI C1 exhibits a very narrow distribution spectrum of the intensity values of the pixels. The maximal variability was observed on intensities for the minValues at 16% and decreased to 4% for maxValues and meanValues.

Heterogeneous ROIs C2 and C3 exhibited a close histogram distribution of intensity values of pixels, with a range close to those of human tissues. Mean values of C2 and C3 were close (<6%) despite differences in the minimal and standard pixel values. The maximal variabilities were also observed for the minValues at 19% and were less than 1% for the other values. The fluctuation of the mean value of C1 is larger than in C2 and C3, this may be explained by the fact that the stochastic noise effect may be more marked in a homogeneous area than in heterogeneous areas.

The CVs of texture indices for the three ROI pooled from the eight identical CT acquisitions are shown in Table 3. Values obtained from the homogeneous insert C1 were more variable than the C2 and C3 inserts, with CVs ranging from 1 to 26%. Of note, the high variations of skewness and busyness in CV values could be explained by the variations around 0 of the mean μ values of these indices. In the case of busyness, the strong variability implied a lack of reliability. The maximum variations observed for C2 and C3 were between 0 and 18%.

Figures 3, 4 and 5 were built to assess which indices permitted discrimination of each insert. The three indices determined as unable to discriminate C1 from C2 and C3 were busyness, GLNU from the Grey-Level Size Zone Matrix GLSZM (Figure 3) and SZLGE. The following parameters were identified as unable to discriminate C2 from C3: entropy and energy derived from histogram; homogeneity, energy and correlation (Figure 4) from the co-occurrence matrix; SRE, LRE, LGRE, HGRE, SRLGE, SRHGE, LRLGE, LRHGE, RLNU, RP from the Grey-Level Run Length Matrix; and SZE, LZE, LGZE, HGZE, SZHGE, LZLGE, LZHGE, GLNU, ZP from the Grey-Level Size Zone Matrix.

From the repeatability study, eight indices were selected because they allow the discrimination of the three textures between them. These indices are skewness (figure 5), kurtosis derived from histogram, contrast and entropy from the co-occurrence matrix, dissimilarity and contrast from the Neighborhood Grey Level difference matrix, SZE and ZLNU from the Grey-Level Size Zone Matrix.

All eight indices were associated with highly significant p-values (<0.0002; Mann-Whitney test) when comparing C1 to C2, C2 to C3, and C1 to C3, as there was no overlap between the distributions for any of the eight parameters.

Finally, redundant indices were identified with a 2x2 correlation matrix (Table 4, see supplemental digital content) and a principal component analysis derived from the 39 supplementary acquisitions described in Table 2. Based on the analysis of the 2x2 correlation matrix, skewness and kurtosis were independent of the six other indices but were inter-correlated for C2 and C3, and not for C1. The other six indices correlated to various degrees: four indices (contrast of the co-occurrence matrix, contrast of the Neighborhood Grey Level difference matrix, entropy, and dissimilarity - were all associated with correlation coefficients >0.8, even for C1.

SZE was highly correlated with these four indices for C2 and C3, but to a lesser extent for C1. ZLNU was highly correlated with the five other indices for C1, but to a lesser extent for C2 and C3.

Principal component analyses performed separately for each phantom pattern (C1, C2, C3) confirmed the 2x2 correlation analyses (Figures 6 A, B and C: plots of the 1st and 2nd principal components, see supplemental digital content). The patterns of the three graphs were very similar. Six indices (dissimilarity, contrast, contrast.1, entropy, ZLNU and SZE) contribute almost equally (around 15%) to the first principal component. A “size effect” was observed, i.e. highly positively correlated variables are on the same side of the axis, with these six indices on the right side of the first axis. The second principal factor is determined by skewness and kurtosis whose cumulative contribution to the second axis is above 85% for C1, C2 and C3. The figures show the percentage of variance due to each component and the contribution of each index.

Discussion

The use of texture analyses in medical imaging is an attractive option, potentially allowing the development of a novel form of disease "biomarkers". Several teams worldwide have developed software for texture analysis and have published *in vivo* data in widely varying fields, supporting the potential for major advances in our understanding of diseases[12].

It is currently very difficult to draw a coherent position or recommendations from this abundant literature especially in order to select the proper texture indices that could be relevant in clinical practice[2].In particular, the reproducibility of the calculation of textural features is currently a subject of debate. Our study indicates that many textural indices show a lack of reproducibility, even under the same experimental conditions. We can hypothesize that the variations we observed are due to the complexity of helical CT principles which involve detectors properties, filtered back projection algorithm and intrinsic noise. We also observed higher fluctuations within the homogenous area C1 than within C2 and C3. This could be explained by the use of an absolute discretization pixel intensity by steps of 10 HU in our experimental conditions which would have a greater influence on the smaller range of values of C1 (around 40 HU) compared to C2 and C3 (both around 300 HU).

To our knowledge, this is the first study exploring the reproducibility of textural indices from CT images acquired eight times on a single device and a dedicated phantom. Some authors have already worried about this subject in vivo but the analysis of the literature is rendered difficult because the softwares and mathematical definition of each textural index used for textural indices often differ to some extent as no consensus currently exists[12,13].

Fave et al[14] studied the reproducibility of test-retest cone beam CT images of 10 patients explored with 68 textural features (histogram, co-occurrence matrix, run length matrix, NGLDM) computed with Ibex software. They excluded 23 features because they were not reproducible on the test-retest images; unfortunately, with the exception of skewness derived from a histogram and contrast derived from the co-occurrence matrix, it is very difficult to compare their data with our study as we did not extract identical indices.

Hunter et al[5] studied the reproducibility of test-retest of unenhanced CT images of 56 patients with lung tumors using three different devices (15, 16 and 25 patients for each device) explored with 328 textural features calculated by Ibex software (including histogram, co-occurrence matrix and Grey Level Run Length Matrix). They found a high concordance correlation coefficient (>0.90) between the two acquisitions for 61.0% to 94.5% of the features, and a high concordance correlation coefficient for 138 image features when interpolating the reproducible features of each device. Of them, they found 23 non-redundant features, including kurtosis, skewness derived from the histogram, and entropy derived from the co-occurrence matrix according to our results. Balagurunathan et al[15] studied the

reproducibility of 219 3D-textural indices derived from test-retest unenhanced CT scans of 32 patients with lung tumors and confirmed the robustness of contrast derived from the co-occurrence matrix.

Moreover, this is the first study to assess the usefulness of indices for characterizing CT medical images on a dedicated phantom. Our study shows that some textural indices are unable to discriminate homogeneous from heterogeneous features, likely rendering these indices unfeasible for use in clinical evaluation. Taking into account that our experimental conditions did not evaluate the influence of the size of the ROI, some indices are unable to discriminate two heterogeneous ROIs.

Our study has a number of limitations. We studied the indices extracted from the histogram and four matrices (co-occurrence, Grey Level Run Length Matrix, Neighborhood Grey Level difference matrix and Grey-Level Size Zone Matrix) from a single software (LIFEx), which is not as exhaustive as some published research,[4,14,15] however our findings are in agreement with data published by other groups in terms of the fact that only a few textural indices are likely to be sufficiently robust to be handled in clinical practice. In addition, we did not evaluate the influence of the size of the ROI. In the study of Fave et al[14] the authors excluded the following features: energy of the histogram, correlation of the co-occurrence matrix, LRHGE, SRLGE of the RLM matrix, busyness and contrast of the NGLDM matrix because their values were found to be volume-dependent.

We did not evaluate the influence of 3D vs 2D evaluation, but 2D images are more robust to variability than 3D images in patient studies.[16] In clinical routine and especially in oncology, tumor images are often ill-defined and subject to approximation in their segmentation, hence tumor segmentation is an important step to address before analysis of textural features is possible. Finally, we chose a discretization of 200 grey levels, following results from other groups with LIFEx software in PET-CT imaging.[9,17]

In conclusion, our preliminary study highlighted a need for close attention when using texture analysis as a tool to characterize CT images because changes in quantification may be due to internal variability rather than due to actual physio-pathological effects. Some textural indices appear to be broadly influenced by noise, whereas 8 out of 34 indices such as skewness, kurtosis derived from histogram, contrast and entropy from the co-occurrence matrix, dissimilarity and contrast from the Neighborhood Grey Level difference matrix

NGLDM and SZE and ZLNU from the Grey-Level Size Zone Matrix GLSZM appear to be sufficiently reliable and capable of discriminating close textures on CT images. Further research should be carried out to confirm and extend these results.

Disclosure of conflicts of Interest: The authors have no relevant conflicts of interest to disclose

References

1- Castellano G, Bonilha L, Li LM, et al. Texture analysis of medical images. Clin Radiol. 2004;59:1061–1069.

2- Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016;278(2):563-577.

3- Limkin EJ, Sun R, Dercle L et al. Promises and challenges for the implementation of computationnal medical imaging (radiomics) in oncology. Ann Oncol. 2017;28(6):1191-1206.

4- Fave X, Cook M, Frederick A et al. Preliminary investigation into sources of uncertainty in quantitative imaging features. Comput Medical Imaging Graph. 2015;44:54–61.

5- Hunter L, Krafft S, Stingo F, et al. High quality machine-robust image features: identification in non-small cell lung cancer computed tomography images. Med Phys. 2013;40(12):121916.

6- Zhao B, Tan Y, Tsai WY, et al. Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study. Transl Oncol. 2014;7(1):88-93.

7- Mackin D, Fave X, Zhang L, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. Invest Radiol. 2015;50(11):757-765.

8- Orlhac F, Soussan M, Chouahnia K, et al. 18F-FDG PET-Derived Textural Indices Reflect Tissue-Specific Uptake Pattern in Non-Small Cell Lung Cancer. PLoS One. 2015;10(12):e0145063.

9- Nioche C, Orlhac F, Boughdad S, et al. A freeware for tumor heterogeneity characterization in PET, SPECT, CT, MRI and US to accelerate advances in radiomics. J Nucl Med 2017 58:1316

10- O'Connor JP, Aboagye EO, Adams JE. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol. 2017 Mar;14(3):169-186

11- Abdi H, Williams LJ. Principal component analysis. Wiley Interdisciplinary Reviews: Computational Statistics, 2010; 2, 433-459.

12- Bashir U, Siddique MM, Mclean E, et al. Imaging Heterogeneity in Lung Cancer:

Techniques, Applications, and Challenges. AJR Am J Roentgenol. 2016;207(3): 534-43.
13- Buvat I, Orlhac F, Soussan M.. Tumor Texture Analysis in PET: Where Do We Stand? J Nucl Med. 2015;56(11):1642-1644.
14- Fave X, Mackin D, Yang J et al. Can radiomic features be reproducibly measured from CBCT images for patients with non small cell lung cancer? Med Phy. 2015;42(12):6784-6797
15- Balagurunathan Y, Gu Y, Wang H, et al. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. 2014;7(1):72-87.
16- Zhao B, Tan Y, Bell DJ, et al. Exploring intra- and inter-reader variability in uni-dimensional, bi-dimensional, and volumetric measurements of solid tumors on CT scans reconstructed at different slice intervals. Eur J Radiol. 2013;82(6):959-968.
17- Orlhac F, Soussan M, Chouahnia K, et al. 18F-FDG PET-Derived Textural Indices Reflect Tissue-Specific Uptake Pattern in Non-Small Cell Lung Cancer. PLoS One. 2015;10(12): e0145063
18- Haralick R, Shanmugam K, Dinstein I. Textural Features for Image Classification. In IEEE Transactions on Systems, Man, and Cybernetics. Volume SMC-3 (6); 1973: 610-621.
19- Galloway M,. Texture analysis using gray level run lengths. Computer Graphics and Image Processing. Volume 4, Issue 2; 1975:172-179.
20- Amadasun M, King R. Textural features corresponding to textural properties. In IEEE Transactions on Systems, Man, and Cybernetics. *V*olume 19 (5); 1989:1264-1274.
21- Thibault G. Texture Indexes and Gray Level Size Zone Matrix Application to Cell Nuclei Classification, INSERM UMR 910, Medical Genetic and Functional Genomic, Medical

Figure 1: C2 and C3 inserts compacted into two syringes adapted to precisely fill the holes of the CIRS® 062M phantom. Resulting histogram of the 3 ROI.

Table 1: Design of the 40 different CT acquisitions performed for the redundancy analysis. The first CT acquisition corresponds to the acquisition repeated eight times for the reproducibility analysis.

Figure 2: Designed phantom with different inserts and resulting CT image of one slice of the phantom: 2D identical regions of interest were drawn on each of the three inserts: C1 for the homogeneous insert (red), C2 for the insert made of Ecoflex® + carbon1 (blue) and C3 for the insert made of Ecoflex® + carbon2 (yellow).

Table 2: Descriptive values of each ROI in Hounsfield units for the 8 identical CT acquisitions (1 to 8).

Table 3: Coefficient of variation (CV) of the values of each index calculated from the eight identical CT acquisitions for C1, C2 and C3 areas.

Figure 3: Dot plot of the close distribution of values of the calculation of GLNU index from the Grey-Level Size Zone Matrix GLSZM of C1, C2 and C3.

Figure 4: Dot plot of the close distribution of values of the calculation of the correlation index between C2 and C3.

Figure 5: Dot plot of the difference of the distribution of values of the calculation of skewness between the three ROI.

Table 1: Design of the 40 different CT acquisitions performed for the redundancy analysis. The first CT acquisition corresponds to the acquisition repeated eight times for the reproducibility analysis.

Acquisition	kVp	mAs	slice thickness (mm)	pitch	rotation time (s)	DFOV (mm)	ASIR (%)	filter
base	120	150	1.25	1.375	0.6	360	0	standard
2	80	150	1.25	1.375	0.6	360	0	standard
3	100	150	1.25	1.375	0.6	360	0	standard
4	140	150	1.25	1.375	0.6	360	0	standard
5	120	100	1.25	1.375	0.6	360	0	standard
6	120	200	1.25	1.375	0.6	360	0	standard
7	120	250	1.25	1.375	0.6	360	0	standard
8	120	300	1.25	1.375	0.6	360	0	standard
9	120	350	1.25	1.375	0.6	360	0	standard
10	120	400	1.25	1.375	0.6	360	0	standard
11	120	500	1.25	1.375	0.6	360	0	standard
12	120	150	0.625	1.375	0.6	360	0	standard
13	120	150	2.5	1.375	0.6	360	0	standard
14	120	150	3.75	1.375	0.6	360	0	standard
15	120	150	5	1.375	0.6	360	0	standard
16	120	150	1.25	0.516	0.6	360	0	standard
17	120	150	1.25	0.984	0.6	360	0	standard
18	120	150	1.25	1.375	0.4	360	0	standard
19	120	150	1.25	1.375	0.5	360	0	standard
20	120	150	1.25	1.375	0.7	360	0	standard
21	120	150	1.25	1.375	0.8	360	0	standard
22	120	150	1.25	1.375	0.9	360	0	standard
23	120	150	1.25	1.375	1	360	0	standard
24	120	150	1.25	1.375	0.6	160	0	standard
25	120	150	1.25	1.375	0.6	260	0	standard
26	120	150	1.25	1.375	0.6	460	0	standard
27	120	150	1.25	1.375	0.6	360	50	standard
28	120	150	1.25	1.375	0.6	360	10	standard
29	120	150	1.25	1.375	0.6	360	20	standard
30	120	150	1.25	1.375	0.6	360	30	standard
31	120	150	1.25	1.375	0.6	360	40	standard
32	120	150	1.25	1.375	0.6	360	60	standard
33	120	150	1.25	1.375	0.6	360	70	standard

34	120	150	1.25	1.375	0.6	360	80	standard
35	120	150	1.25	1.375	0.6	360	90	standard
36	120	150	1.25	1.375	0.6	360	100	standard
37	120	150	1.25	1.375	0.6	360	0	soft
38	120	150	1.25	1.375	0.6	360	0	detail
39	120	150	1.25	1.375	0.6	360	0	bone
40	120	150	1.25	1.375	0.6	360	0	lung

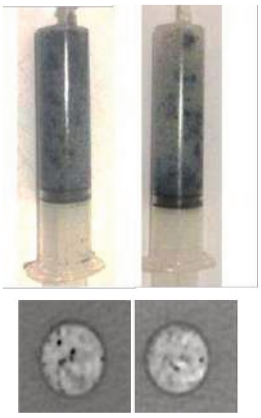
Table 2: Descriptive values of each ROI in Hounsfield units for the 8 identical CT acquisitions (1 to 8).

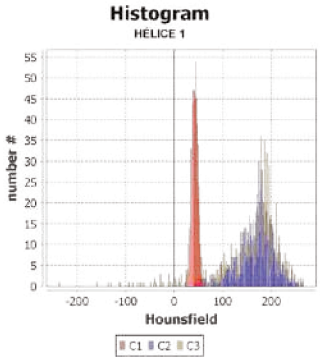
	CT acquisition							
	1	2	3	4	5	6	7	8
Name of ROI : C1								
minValue	25.0	21.0	20.0	14.0	23.0	22.0	20.0	24.0
maxValue	58.0	63.0	60.0	62.0	58.0	58.0	56.0	60.0
stdValue	6.0	6.3	6.2	6.3	6.4	5.9	5.7	5.9
meanValue	41.5	42.5	38.8	38.8	41.4	39.8	38.4	41.5
Name of ROI : C2								
minValue	-39.0	-32.0	-52.0	-52.0	-57.0	-45.0	-52.0	-62.0
maxValue	267.0	278.0	272.0	267.0	267.0	268.0	265.0	273.0
stdValue	41.9	43.0	42.9	42.8	42.6	42.5	42.3	43.2
meanValue	166.2	164.9	167.1	167.0	165.6	165.9	166.4	165.0
Name of ROI : C3								
minValue	-238.0	-237.0	-247.0	-238.0	-218.0	-246.0	-244.0	-246.0
maxValue	265.0	265.0	271.0	258.0	265.0	266.0	264.0	267.0
stdValue	65.5	64.6	66.8	66.4	63.9	65.7	66.1	66.1
meanValue	156.2	157.7	156.5	156.5	157.5	155.6	156.8	156.4

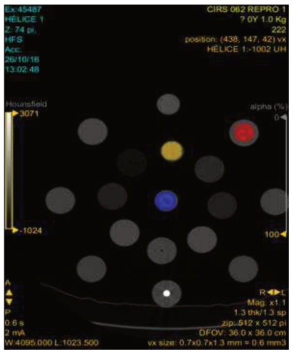
Table 3: Coefficient of variation (CV) of the values of each index calculated from the eight identical CT acquisitions for C1, C2 and C3 areas.

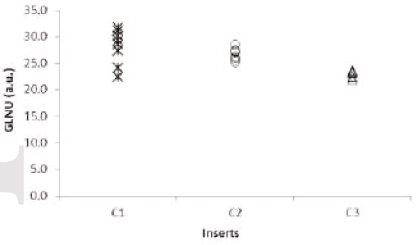
LIFEx	CV	CV	CV
Name of ROI	C1	C2	C3
minValue	15.05	-18.65	-3.72
meanValue	3.60	0.47	0.40
stdValue	3.49	0.90	1.38
maxValue	3.66	1.51	1.27
SkewnessH	-375.13	-4.44	-1.80
KurtosisH	5.11	3.37	2.28
EntropyH	1.00	0.54	0.27
EnergyH	3.17	3.20	1.45
Homogeneity	1.12	0.88	0.80
Energy	9.63	4.02	2.43
Contrast (co-occurrence)	6.14	2.49	3.00
Correlation	8.50	0.80	0.38
Entropy	3.50	0.35	0.34
Dissimilarity	3.92	0.84	1.30
SRE	1.85	0.42	0.44
LRE	10.70	2.29	2.25
LGRE	0.30	0.17	0.18
HGRE	0.30	0.17	0.14
SRLGE	2.01	0.38	0.41
SRHGE	1.73	0.54	0.54
LRLGE	10.82	2.33	2.19
LRHGE	10.58	2.25	2.30
GLNU	2.34	1.22	2.48
RLNU	4.23	1.59	1.54
RP	2.53	0.67	0.59

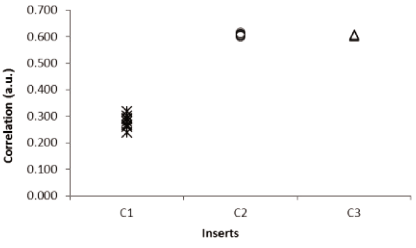
Coarseness	4.46	2.21	1.43
Contrast (NGLDM)	26.22	3.98	5.99
Busyness	-744.65	-4039.39	-202.31
SZE	8.37	1.68	1.28
LZE	12.21	11.87	17.87
LGZE	0.22	0.28	0.31
HGZE	0.22	0.27	0.25
SZLGE	8.29	1.50	1.35
SZHGE	8.46	1.91	1.44
LZLGE	12.19	11.77	17.51
LZHGE	12.24	11.94	18.14
GLNUz	11.10	3.73	2.65
ZLNU	14.30	5.08	3.59
ZP	6.72	2.56	1.41

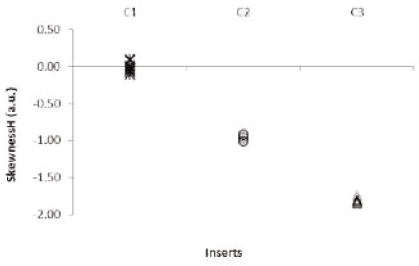


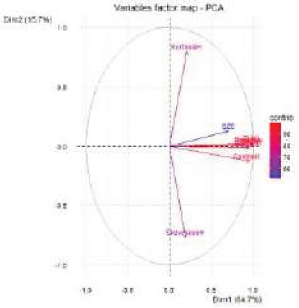


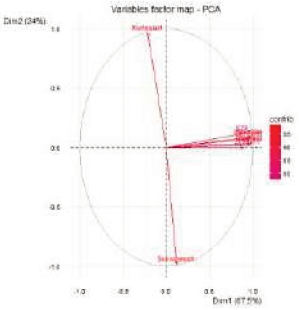


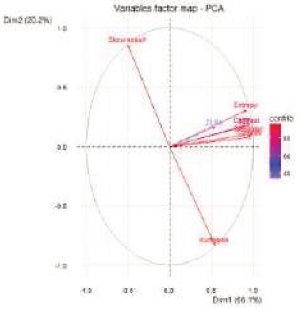












diagnostics

MDPI

Review

CT Texture Analysis Challenges: Influence of Acquisition and Reconstruction Parameters: A Comprehensive Review

Mathilde Espinasse [1,2], Stéphanie Pitre-Champagnat [1], Benoit Charmettant [1], Francois Bidault [1,3], Andreas Volk [1], Corinne Balleyguier [1,3], Nathalie Lassau [1,4,*] and Caroline Caramella [1,3]

1 IR4M-UMR8081, CNRS, Univ Paris Sud, University Paris Saclay, Rue Ampère, 91405 Orsay CEDEX, France; mathilde.espinasse@gustaveroussy.fr (M.E.); stephanie.champagnat-pitre@gustaveroussy.fr (S.P.-C.); benoit.charmettant@gustaveroussy.fr (B.C.); francois.bidault@gustaveroussy.fr (F.B.); andreas.volk@gustaveroussy.fr (A.V.); corinne.balleyguier@gustaveroussy.fr (C.B.); caroline.caramella@gustaveroussy.fr (C.C.)

2 École Normale Supérieure Paris-Saclay, 61 avenue du Président Wilson, 94235 Cachan CEDEX, France

3 Department of Imaging, Gustave Roussy Cancer Campus, 114 rue Edouard Vaillant, 94805 Villejuif CEDEX, France

4 Research Department, Gustave Roussy Cancer Campus, 114 avenue Edouard Vaillant, 94805 Villejuif, France

* Correspondence: Nathalie.lassau@gustaveroussy.fr; Tel.: +3-31-4211-6014

Received: 7 April 2020; Accepted: 24 April 2020; Published: 28 April 2020

Abstract: Texture analysis in medical imaging is a promising tool that is designed to improve the characterization of abnormal images from patients, to ultimately serve as a predictive or prognostic biomarker. However, the nature of image acquisition itself implies variability in each pixel/voxel value that could jeopardize the usefulness of texture analysis in the medical field. In this review, a search was performed to identify current published data for computed tomography (CT) texture reproducibility and variability. On the basis of this analysis, the critical steps were identified with a view of using texture analysis as a reliable tool in medical imaging. The need to specify the CT scanners used and the associated parameters in published studies is highlighted. Harmonizing acquisition parameters between studies is a crucial step for future texture analysis.

Keywords: radiomics; texture analysis; computed tomography; acquisition parameters

1. Introduction

Computed tomographic (CT) images are routinely used for the diagnosis and follow-up of patients. These images represent a huge amount of numerical data, which are both transferable and storable.

Recently, the temptation of using these data as potential biomarkers for a large range of clinical questions, especially the prediction of response and prognosis, has modified the way researchers are bethinking imaging. Texture analysis is based on many well described mathematical approaches such as first and second order texture calculation, combined with the application of different mathematical filters, which can lead to up to 1000 texture features for one image. Translating the medical imaging numerical datasets into quantitative figures, and thus authorizing statistical comparisons, is known as radiomics [1]. Numerous publications have already studied texture as a diagnosis or prognosis biomarker, but results are very heterogeneous and confusing [2,3]. Indeed, this new field of research comes with a new language, new concepts and a lack of international standardization.

Although efforts from several international communities, such as the Quantitative Imaging Biomarkers Alliance (QIBA) [4], the Quantitative Imaging Network (QIN) [5] or the Image Biomarker Standardisation Initiative, exist [6], this review is aimed at helping radiologists to better understand the

current scientific data specifically based on CT texture variability, and to point out the many challenges texture applied to medical imaging has to face before it could become a reliable biomarker.

2. Methods

For the purpose of this work, articles were selected through a PubMed search, using the keywords ("computed tomography" OR "CT") AND ("texture" OR "radiomics"). A cut-off of 5 February 2018 was used and 1143 articles were identified.

Studies dealing with the prognosis, diagnosis and responses to treatment of patients, focusing on other imaging techniques than CT or using only first order texture features (histograms) were discarded. Articles retained included patients or phantom studies focusing on the impact of the acquisition and processing parameters on the variability, reproducibility and repeatability of texture features. Finally, 20 articles were included in the study. Table 1 summarises the main characteristics of the selected articles.

Table 1. Characteristics of selected articles. CCR refers to the credence cartridge radiomics phantom, RIDER to the Reference Image Database to Evaluate Therapy Response and NSCLC to non-small-cell lung carcinoma. * Fave et al. do not indicate the number of CT. † The number is not stated by the authors but the patients come from another study, which included 107 patients.

Reference	Phantom	Patients	Number of CT Devices	Number of Patients	Software	Parameters Studied
Al-Kadi 2009 [7]	No	Lung	2	67	In-house	Repeatability
Balagurunathan 2014 [8]	No	RIDER	2	32	In-house	2D/3D
Berenguer 2018 [9]	Pelvic + CCR copy	No	5	NA	IBEX	Repeatability and redundancy, various acquisition parameters
Buch [10]	In-house	No	1	NA	LIFEx	Tube voltage, current, slice thickness
Caramella 2018 [11]	In-house	No	2	NA	In-house	Repeatability
Fave 2015a [12]	No	NSCLC	? *	20	IBEX	Voltage, current, 2D/3D
Fave 2015b [13]	CCR	NSCLC	19	10	IBEX	Repeatability, CT scanner brand
He 2016 [14]	No	Lung	1	240	In-house	contrast enhancement
Kim 2016 [15]	No	Lung nodule	1	42	In-house	Reconstruction algorithm
Larue 2017 [16]	CCR	NSCLC	9	325	In-house	Repeatability, current, slice thickness
Lu 2016 [17]	No	RIDER	1	32	In-house	Slice thickness, filter
Mackin 2015 [18]	CCR	NSCLC	16	20	IBEX	CT scanner brand
Mackin 2017 [19]	No	NSCLC	1	8	IBEX	Pixel size
Mackin 2018 [20]	CCR	NSCLC	2	107 †	IBEX	Current
Mahmood 2017 [21]	Lung	No	3	NA	IBEX	Filter, CT scanner brand
Midya 2018 [22]	Uniform + anthropomorphic	Abdominal scan	1	1	In-house	Current, reconstruction algorithm
Shafiq-ul-Hassan 2017 [23]	CCR	No	8	NA	In-house	Slice thickness, pixel size
Solomon 2016 [24]	No	Lung, liver, kidney	1	20	In-house	Reconstruction algorithm
Yang 2015 [25]	No	Lung	1	8	IBEX	Contrast enhancement
Zhao 2014 [26]	Thorax	No	1	NA	In-house	Slice thickness, filter

The texture features studied in the selected articles included first order indices and higher order features. First order features are parameters extracted from the histogram of the distribution of the values of pixels. Second order features come from the matrices describing the spatial relations between pixels: grey level co-occurrence matrix (GLCM), grey level run length matrix (GLRLM), neighbourhood grey level difference matrix (NGLDM), grey level size zone matrix (GLSZM), grey level zone length matrix (GLZLM), etc. Texture features can also come from fractal or wavelet techniques or Gaussian Markov random fields.

Articles often dealt with both patients and phantoms, or with more than one parameter; this implies that the same article can be found in numerous parts of this paper.

3. Results

3.1. Texture Processing

3.1.1. Software

The choice of software is linked with the method selected. Numerous texture analysis software exist, for example LIFEx [27], IBEX [28], Pyradiomics [29] or MaZda [30]. These software tend to generate a large number of texture features, of which many are common to all software, but not all studies use the same descriptors which makes it difficult to compare the results. Furthermore, it is important to note that sometimes the same name of texture feature can cover different computation methods or different feature names can actually represent the same quantity as described by Buvat et al. [31]. Numerous teams also developed their own in-house software. In order to deal with this issue, developers need to align with current recommendations offered by the Image Biomarker Standardisation Initiative [6] which provides standardized nomenclature and definitions, a standardized image processing workflow and implements guidelines for conducting radiomics studies.

3.1.2. Dimensionality

Features can be calculated in 2, 2.5 or 3 dimensions. These different computation methods give different results, as highlighted by Balagurunathan et al. [8]. They analyse 219 3D texture features and 110 2D with scans from 32 patients from the RIDER database, and they conclude that 3D features better describe the volume but 2D features are more easily interpreted. Fave et al. [12] computed 23 texture features both in 3D and in 2D on the largest cross-sectional slice of patients' CT scans: 8 varied significantly between 2D and 3D, but 14 were significantly correlated between 2D and 3D with a Spearman correlation coefficient over 0.85. They concluded that the majority of 2D and 3D features translate the same heterogeneity, but that these two computation methods cannot be mixed as the numerical results are different. Despite the fact that the values derived from an analysis of the largest cross-sectional slice seem to be an effective substitute for a whole tumour analysis, they recommend the whole tumour analysis whenever possible to avoid any bias induced by the choice of the slice.

3.2. Texture Repeatability

3.2.1. Intra-Patients Repeatability

Balagurunathan et al. [8] studied the RIDER database which contains repeat CT scans of lung cancer patients performed with the same CT scanner with a 15-minute interval and reported 48 reproducible features (concordance correlation coefficient (CCC) > 0.9) out of 219. Fave et al. [13] extracted texture features from the CBCT scans of 10 patients and excluded 23 features out of 68 as non-reproducible (CCC $\leq$ 0.9). They also demonstrated with a dynamic-motion thorax phantom, that out of 68 features, 12 are reproducible with a 4 mm movement, and only 3 with a 6 to 8 mm movement.

3.2.2. Phantoms

The majority of repeatability studies are performed on phantoms. Caramella et al. [11] conducted eight consecutive CT scans on the same in-house phantom with the same CT scanner using the same parameters. They extracted 34 features with LIFEx and kept only 8 as reproducible. This emphasizes a lack of experimental reproducibility under the same experimental conditions, which might be of greater concerns in vivo. Berenguer et al. [9] also tested the reproducibility and redundancy of texture features computed with IBEX with test–retest, intra-CT and inter-CT analyses. The CT acquisition parameters remained identical in the test–retest and inter-CT analyses. They found 161 features out of 177 as reproducible, but the redundancy study concluded that the 177 studied features could be summarized by 10 of them. The study acknowledged that multicentre reproducibility is of a great challenge but it can be minimized using rigorous acquisition protocols.

This partial non-reproducibility of texture features may be related to stochastic noise. Al-Kadi et al. [7] studied the impact of various distributional noise on 74 texture features from 7 different computation methods (including different matrices, wavelets and fractal dimension). The enhanced and unenhanced CT scans of 67 patients, taken on 2 different CT scans of lungs showing tumours at different stages were used. The study concluded that they were affected by noise, but differently for each feature. The features with the highest characterization power were the least affected by noise. They showed that adaptive filtering can help reduce subtle noise.

3.3. Intrinsic CT Parameters

3.3.1. CT Scanner Brand

Each manufacturer uses its own X-ray tube, detectors, reconstruction and post-processing algorithms to build the image.

Mackin et al. [18] and Fave et al. [13] studied the influence of CT scanners on texture features with the same in-house phantom containing cartridges of different materials (later referred to as the CCR phantom). Mackin et al. compared the interscanner and interpatient variabilities of texture features on 16 CT scanners from four different manufacturers, each with its own standard acquisition protocol. They found that interscanner variability depends on the feature under consideration and the material of the region of interest, but showed that the interscanner are of the same order of magnitude than the interpatient variabilities. Fave et al. scanned the CCR phantom on two cone beam CT scanners, using different acquisition parameters. They got a good reproducibility of features when comparing CT scans acquired from the same manufacturer, whereas using different protocols limited the reproducibility, and comparing the different manufacturers completely withdrew reproducibility.

Larue et al. [16] studied a modified CCR phantom with textured inserts on nine different CT scanners with fixed voltage, pitch and computed tomography dose index and extracted 114 texture features with their in-house software. The distribution of the features' values was different, implying that the variability was related to the CT scanner brand.

Buch et al. [10] also explored an in-house phantom made out of cereal and mayonnaise on two different CT (same brand but different number of detectors 16b and 64b) and demonstrated a significant difference in the computation of the histogram and GLCM features.

Mahmood et al. [21] used an anthropomorphic lung phantom with shredded rubber and sycamore wood inserts to perform acquisitions on machines from three manufacturers, with a constant voxel size, kVp, pitch and slice thickness. None of the 27 texture features computed with IBEX passed the reproducibility criteria.

3.3.2. Reconstruction Algorithm

Solomon et al. [24] studied three reconstruction algorithms: filtered back projection (FBP), adaptive statistical iterative reconstruction (ASIR) and model-based iterative reconstruction (MBIR). Twenty patients' diverse conditions were scanned on a single CT scanner, and 23 features were extracted with the three algorithms. Compared with using reference conditions using an FBP reconstruction algorithm at a high dose, between 1 and 3 features for ASIR and between 9 and 11 features for MBIR were affected by a change in the reconstruction algorithm, depending on the organ involved. This suggests a significant impact of the reconstruction algorithm on the texture analysis.

Midya et al. [22] also studied the role of ASIR. They extracted 248 features from CT scans performed on a uniform water phantom, an anthropomorphic phantom and one patient. They observed that an increase in the percentage of ASIR compared to FBP alone increased the blurring of the image and decreased the number of comparable texture features, meaning different ASIR levels could not be mixed in the same study.

Kim et al. [15] studied the influence on texture features of choosing FBP or sinogram affirmed iterative reconstruction (SAFIRE). They studied lung nodules in 42 patients and extracted 15 features.

They showed that among those features, five first order tumour intensity features and four co-occurrence GLCM-based features were significantly affected by the choice of reconstruction algorithm. They however noted that the inter-reader variability induced by the segmentation of the region of interest (ROI) was significantly higher than the one induced by the reconstruction algorithm for nine features but for entropy, homogeneity and the four GLCM-based features, the inter-reconstruction algorithm variability was greater.

3.4. Acquisition Parameters

3.4.1. Tube Voltage

Tube voltage sets the number and energy of produced photons and is fixed prior to the acquisition. Fave et al. [13] studied the influence of a change in kVp on 23 features extracted with IBEX from the CBCT scans of 20 patients taken at 120 kVp and 300 mA. Through a simulation algorithm, they explored the effects of 80 kVp, 100 kVp and 140kVp and showed that the intrapatient variability due to a change in voltage was always inferior to the interpatient variability. Buch et al. [10] showed the same results on their in-house phantom when applying 80 kv to 140 kv. Indeed, they studied a set of 42 texture features derived from CT scan images using a custom software. They witnessed no significant statistical variation for any of the included features.

3.4.2. Tube Current

Tube current sets the number of photons and is fixed prior to the acquisition.

Midya et al. [22] performed CT scans with varying tube currents (50 to 500 mA) on a uniform water phantom. The 248 texture features extracted with the in-house software varied with the changes in tube current, particularly when dealing with low tube currents. Mackin et al. [20] extracted 48 texture features from homogeneous and heterogeneous regions of the CT scans of the CCR phantom with a tube current covering 25 to 300 mA. They also concluded that texture features extracted from the homogeneous regions were very dependent on the current value, and had a higher variability than the interpatient variability, while texture features extracted from the heterogeneous regions were less sensible to the current variation. Fave et al. [13] also modified the images obtained on 20 patients (see Section 3.4.1) to simulate different values of tube current, from 100 to 300 mA. They observed that 10 out of 23 texture features had a lower intrapatient variability due to current change than the interpatient variability, but that for 13 out of 23 features, the intrapatient and interpatient variability were of the same order of magnitude.

On the contrary, Larue et al. [16] scanned the same CCR phantom with nine different CT scanners and different tube currents. The analysis of 114 texture features extracted from a heterogeneous region with in-house software did not reveal a clear influence of the tube current on the texture features but acknowledged that such an investigation deserved to be performed on a larger dataset. Buch et al. [10] also showed no influence of the tube current variation, ranging only from 80 to 120 mAs.

3.4.3. Slice Thickness, Pixel Size

Slice thickness and pixel size both determine the voxel size, which in turn determines the spatial resolution of the image.

Shafiq-ul-Hassan et al. [23], studied the effect of the pixel size and slice thickness on 213 texture features on the CCR phantom. They acquired images with different slice thicknesses and with different fields of view. They subsequently resampled the voxel size to $1 \times 1 \times 2$ mm^3, and compared the resampled and non-resampled images' features: 150 were unaffected by the resampling, 42 were significantly improved and 21 were still variable. Larue et al. [16] confirmed their results: they perform CT scans with slice thicknesses of 1.5 and 3 mm on the same phantom and concluded that a large proportion of 114 texture features were affected by the changes and that variability was reduced after resampling the voxel size to $1 \times 1 \times 3$ mm^3.

Zhao et al. [26], Lu et al. [17] and Buch et al. [10] conducted the same kind of study; Zhao et al. on a thorax phantom, Lu et al. on 32 patients from the RIDER database and Buch et al. on an in-house phantom, and they all concluded that texture features changed significantly with the slice thickness. Mackin et al. [19] studied the impact of pixel size on intrapatient variability. Their study included eight NSCLC patients and they calculated 150 2.5D texture features (texture features calculated slice by slice then combined) and highlighted that most were dependent on pixel size. They then corrected the differences in pixel size by resampling and filtering, and decreased from 80% to 10% the proportion of features with a higher variability due to pixel size rather than interpatient variability.

3.4.4. Filter

Many filters are provided by CT scanner devices and are named differently according to each brand. Zhao et al. [32] and Lu et al. [17] showed the dependence of texture features on the chosen filter, respectively on a thorax phantom and on 32 patients. The images were reconstructed with both lung and standard filters and they concluded that the chosen filter influenced the value of the texture features.

Mahmood et al. [21] studied 27 texture features extracted from the CT scans of a lung phantom. They only focused on two different kind of secondary order features: neighbourhood grey-tone difference matrix (NGTDM) and GLCM extracted with the IBEX radiomics software. The phantom was scanned on three CT scanners and the images were reconstructed with standard/B40f and lung/B60f filters. They found that none of the features were reproducible when the CT scans were taken with the same manufacturer but reconstructed with different filters.

3.5. Contrast Enhancement

Yang et al. [25] studied the dependency of texture features with a time elapse between the injection of the contrast product and the acquisition of the CT scan. They scanned eight patients in two sessions, six times per session, and extracted 122 texture features of lung tumours with IBEX. For seven of their patients, there was no obvious correlation between the time of acquisition and texture features. He et al. [14] extracted 105 texture features using an in-house feature extraction algorithm from 240 CT scans of patients with a lung nodule from both unenhanced and enhanced (25 s after injection) images. They assessed the discriminatory power of each feature using a Mann–Whitney U test in a univariate analysis. They then performed feature selection and dimensionality reduction to build a radiomic signature for each image. Then, they finally analysed the discrimination and classification performance of the radiomic signature and compared the performances for the different sets of CT scans, with or without contrast enhancement. They concluded that UECT gives better results in the discrimination and classification of nodules.

4. Conclusions

This review highlights the variety of effects that changes in acquisition parameters can have on texture features and the difficulty in the interpretation of texture studies. Tube voltage and current appear to have a limited effect on texture features. Tube current was shown to affect heterogeneous regions to a lesser extent than homogeneous ones. Pixel size and slice thickness have a major influence on texture features, highlighting the need for post-processing resampling. The choice of filter also affects texture features and the question of contrast enhanced images is not yet resolved.

The choice of software, the calculating method (2D or 3D) and the type of CT scanner and brand need to be carefully reported in studies. The question of whether it will be possible through harmonization to get comparable results with CT scanners from different manufacturers is not yet resolved. As studies published outside the scope of this study suggest, the manufacturer variability can be reduced by using a controlled protocol [33]. Moreover, variability also occurs between scans taken on the same CT scanners, and thus asks the question of the accountability of stochastic noise.

This study limited itself to the variability induced by machine related parameters (acquisition and reconstruction parameters). It is important to note that many of the articles reviewed emphasized the importance of human induced variability and in particular the influence of the segmentation of the region of interest, which is seen as a major factor of variability.

This review advocates for the need to state as precisely as possible the methodology regarding the CT scanners (brand, acquisition parameters) used and the post-processing (texture software, if in-house software: definition of algorithms). In CT studies, a harmonization of the acquisition parameters is the key to the future of optimal texture analysis

Funding: This research received no external funding.

Acknowledgments: The authors thank MacKenzie for editing.

Conflicts of Interest: The authors of this manuscript declare no relationships with any companies, whose products or services may be related to the subject matter of the article.

References

1. Gillies, R.J.; Kinahan, P.E.; Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **2016**, *278*, 563–577. [CrossRef] [PubMed]
2. Yip, S.S.F.; Aerts, H.J.W.L. Applications and limitations of radiomics. *Phys. Med. Biol.* **2016**, *61*, R150–R166. [CrossRef]
3. Thawani, R.; McLane, M.; Beig, N.; Ghose, S.; Prasanna, P.; Velcheti, V.; Madabhushi, A. Radiomics and radiogenomics in lung cancer: A review for the clinician. *Lung Cancer* **2018**, *115*, 34–41. [CrossRef] [PubMed]
4. Kessler, L.G.; Barnhart, H.X.; Buckler, A.J.; Choudhury, K.R.; Kondratovich, M.V.; Toledano, A.; Guimaraes, A.R.; Filice, R.; Zhang, Z.; Sullivan, D.C.; et al. The emerging science of quantitative imaging biomarkers terminology and definitions for scientific studies and regulatory submissions. *Stat. Methods Med. Res.* **2014**, *24*, 9–26. [CrossRef] [PubMed]
5. Press, R.H.; Shu, H.-K.G.; Shim, H.; Mountz, J.M.; Kurland, B.F.; Wahl, R.L.; Jones, E.F.; Hylton, N.M.; Gerstner, E.R.; Nordstrom, R.J.; et al. The Use of Quantitative Imaging in Radiation Oncology: A Quantitative Imaging Network (QIN) Perspective. *Int. J. Radiat. Oncol. Biol. Phys.* **2018**, *102*, 1219–1235. [CrossRef] [PubMed]
6. Zwanenburg, A.; Vallières, M.; Abdalah, M.A.; Aerts, H.J.W.L.; Andrearczyk, V.; Apte, A.; Ashrafinia, S.; Bakas, S.; Beukinga, R.J.; Boellaard, R.; et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology* **2020**, 191145. [CrossRef]
7. Al-Kadi, O.S. Assessment of texture measures susceptibility to noise in conventional and contrast enhanced computed tomography lung tumour images. *Comput. Med Imaging Graph.* **2010**, *34*, 494–503. [CrossRef]
8. Balagurunathan, Y.; Gu, Y.; Wang, H.; Kumar, V.; Grove, O.; Hawkins, S.; Kim, J.; Goldgof, D.B.; Hall, L.O.; Gatenby, R.A.; et al. Reproducibility and Prognosis of Quantitative Features Extracted from CT Images. *Transl. Oncol.* **2014**, *7*, 72–87. [CrossRef]
9. Berenguer, R.; Pastor-Juan, M.D.R.; Canales-Vázquez, J.; Castro-García, M.; Villas, M.V.; Legorburo, F.M.; Sabater, S. Radiomics of CT Features May Be Nonreproducible and Redundant: Influence of CT Acquisition Parameters. *Radiology* **2018**, *288*, 407–415. [CrossRef]
10. Buch, K.; Li, B.; Qureshi, M.M.; Kuno, H.; Anderson, S.W.; Sakai, O. Quantitative Assessment of Variation in CT Parameters on Texture Features: Pilot Study Using a Nonanatomic Phantom. *Am. J. Neuroradiol.* **2017**, *38*, 981–985. [CrossRef]
11. Caramella, C.; Allorant, A.; Orlhac, F.; Bidault, F.; Asselain, B.; Ammari, S.; Jaranowski, P.; Moussier, A.; Balleyguier, C.; Lassau, N.; et al. Can we trust the calculation of texture indices of CT images? A phantom study. *Med. Phys.* **2018**, *45*, 1529–1536. [CrossRef] [PubMed]
12. Fave, X.; Cook, M.; Frederick, A.; Zhang, L.; Yang, J.; Fried, D.; Stingo, F.; Court, L. Preliminary investigation into sources of uncertainty in quantitative imaging features. *Comput. Med. Imaging Graph.* **2015**, *44*, 54–61. [CrossRef] [PubMed]
13. Fave, X.; Mackin, D.; Yang, J.; Zhang, J.; Fried, D.; Balter, P.; Followill, D.; Gomez, D.; Jones, A.K.; Stingo, F.; et al. Can radiomics features be reproducibly measured from CBCT images for patients with non-small cell lung cancer? *Med. Phys.* **2015**, *42*, 6784–6797. [CrossRef] [PubMed]

14. He, L.; Huang, Y.; Ma, Z.; Liang, C.; Liang, C.; Liu, Z. Effects of contrast-enhancement reconstruction slice thickness and convolution kernel on the diagnostic performance of radiomics signature in solitary pulmonary nodule. *Sci. Rep.* **2016**, *6*, 34921. [CrossRef]
15. Kim, H.; Park, C.M.; Lee, M.; Park, S.J.; Song, Y.S.; Lee, J.H.; Hwang, E.J.; Goo, J.M. Impact of Reconstruction Algorithms on CT Radiomic Features of Pulmonary Tumors: Analysis of Intra- and Inter-Reader Variability and Inter-Reconstruction Algorithm Variability. *PLoS ONE* **2016**, *11*, e0164924. [CrossRef]
16. Larue, R.T.H.M.; van Timmeren, J.E.; de Jong, E.E.C.; Feliciani, G.; Leijenaar, R.T.H.; Schreurs, W.M.J.; Sosef, M.N.; Raat, F.H.P.J.; van der Zande, F.H.R.; Das, M.; et al. Influence of gray level discretization on radiomic feature stability for different CT scanners tube currents and slice thicknesses: A comprehensive phantom study. *Acta Oncol.* **2017**, *56*, 1544–1553. [CrossRef]
17. Lu, L.; Ehmke, R.C.; Schwartz, L.H.; Zhao, B. Assessing Agreement between Radiomic Features Computed for Multiple CT Imaging Settings. *PLoS ONE* **2016**, *11*, e0166550. [CrossRef]
18. Mackin, D.; Fave, X.; Zhang, L.; Fried, D.; Yang, J.; Taylor, B.; Rodriguez-Rivera, E.; Dodge, C.; Jones, A.K.; Court, L. Measuring Computed Tomography Scanner Variability of Radiomics Features. *Investig. Radiol.* **2015**, *50*, 757–765. [CrossRef]
19. Mackin, D.; Fave, X.; Zhang, L.; Yang, J.; Jones, A.K.; Ng, C.S.; Court, L. Harmonizing the pixel size in retrospective computed tomography radiomics studies. *PLoS ONE* **2017**, *12*, e0178524. [CrossRef] [PubMed]
20. Mackin, D.; Ger, R.; Dodge, C.; Fave, X.; Chi, P.-C.; Zhang, L.; Yang, J.; Bache, S.; Dodge, C.; Jones, A.K.; et al. Effect of tube current on computed tomography radiomic features. *Sci. Rep.* **2018**, *8*, 1–10. [CrossRef] [PubMed]
21. Mahmood, U.; Apte, A.P.; Deasy, J.O.; Schmidtlein, C.R.; Shukla-Dave, A. Investigating the Robustness Neighborhood Gray Tone Difference Matrix and Gray Level Co-occurrence Matrix Radiomic Features on Clinical Computed Tomography Systems Using Anthropomorphic Phantoms. *J. Comput. Assist. Tomogr.* **2017**, *41*, 995–1001. [CrossRef] [PubMed]
22. Midya, A.; Chakraborty, J.; Gönen, M.; Do, R.K.G.; Simpson, A.L. Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility. *J. Med. Imaging* **2018**, *5*, 1. [CrossRef] [PubMed]
23. Shafiq-ul-Hassan, M.; Zhang, G.G.; Latifi, K.; Ullah, G.; Hunt, D.C.; Balagurunathan, Y.; Abdalah, M.A.; Schabath, M.B.; Goldgof, D.G.; Mackin, D.; et al. Intrinsic dependencies of CT radiomic features on voxel size and number of gray levels. *Med. Phys.* **2017**, *44*, 1050–1062. [CrossRef]
24. Solomon, J.; Mileto, A.; Nelson, R.C.; Choudhury, K.R.; Samei, E. Quantitative Features of Liver Lesions Lung Nodules, and Renal Stones at MultiDetector Row CT Examinations: Dependency on Radiation Dose and Reconstruction Algorithm. *Radiology* **2016**, *279*, 185–194. [CrossRef]
25. Yang, J.; Zhang, L.; Fave, X.J.; Fried, D.V.; Stingo, F.C.; Ng, C.S.; Court, L.E. Uncertainty analysis of quantitative imaging features extracted from contrast-enhanced CT in lung tumors. *Comput. Med. Imaging Graph.* **2016**, *48*, 1–8. [CrossRef]
26. Zhao, B.; Tan, Y.; Tsai, W.Y.; Schwartz, L.H.; Lu, L. Exploring Variability in CT Characterization of Tumors: A Preliminary Phantom Study. *Transl. Oncol.* **2014**, *7*, 88–93. [CrossRef] [PubMed]
27. Nioche, C.; Orlhac, F.; Boughdad, S.; Reuzé, S.; Goya-Outi, J.; Robert, C.; Pellot-Barakat, C.; Soussan, M.; Frouin, F.; Buvat, I. LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res.* **2018**, *78*, 4786–4789. [CrossRef] [PubMed]
28. Zhang, L.; Fried, D.V.; Fave, X.J.; Hunter, L.A.; Yang, J.; Court, L.E. ibex: An open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys.* **2015**, *42*, 1341–1353. [CrossRef]
29. Van Griethuysen, J.J.M.; Fedorov, A.; Parmar, C.; Hosny, A.; Aucoin, N.; Narayan, V.; Beets-Tan, R.G.H.; Fillion-Robin, J.-C.; Pieper, S.; Aerts, H.J.W.L. Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Res.* **2017**, *77*, e104–e107. [CrossRef]
30. Szczypiński, P.M.; Strzelecki, M.; Materka, A.; Klepaczko, A. MaZdaA software package for image texture analysis. *Comput. Methods Programs Biomed.* **2009**, *94*, 66–76.
31. Buvat, I.; Orlhac, F.; Soussan, M. Tumor Texture Analysis in PET: Where Do We Stand? *J. Nuclear Med.* **2015**, *56*, 1642–1644. [CrossRef] [PubMed]

32. Zhao, B.; Tan, Y.; Tsai, W.-Y.; Qi, J.; Xie, C.; Lu, L.; Schwartz, L.H. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci. Rep.* **2016**, *6*, 1–7. [CrossRef] [PubMed]
33. Ger, R.B.; Zhou, S.; Chi, P.-C.M.; Lee, H.J.; Layman, R.R.; Jones, A.K.; Goff, D.L.; Fuller, C.D.; Howell, R.M.; Li, H.; et al. Comprehensive Investigation on Controlling for CT Imaging Variabilities in Radiomics Studies. *Sci. Rep.* **2018**, *8*, 1–14. [CrossRef] [PubMed]

iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics

Lesley Seymour, Jan Bogaerts, Andrea Perrone, Robert Ford, Lawrence H Schwartz, Sumithra Mandrekar, Nancy U Lin, Saskia Litière, Janet Dancey, Alice Chen, F Stephen Hodi, Patrick Therasse, Otto S Hoekstra, Lalitha K Shankar, Jedd D Wolchok, Marcus Ballinger, Caroline Caramella, Elisabeth G E de Vries, on behalf of the RECIST working group

Tumours respond differently to immunotherapies compared with chemotherapeutic drugs, raising questions about the assessment of changes in tumour burden—a mainstay of evaluation of cancer therapeutics that provides key information about objective response and disease progression. A consensus guideline—iRECIST—was developed by the RECIST working group for the use of modified Response Evaluation Criteria in Solid Tumours (RECIST version 1.1) in cancer immunotherapy trials, to ensure consistent design and data collection, facilitate the ongoing collection of trial data, and ultimate validation of the guideline. This guideline describes a standard approach to solid tumour measurements and definitions for objective change in tumour size for use in trials in which an immunotherapy is used. Additionally, it defines the minimum datapoints required from future trials and those currently in development to facilitate the compilation of a data warehouse to use to later validate iRECIST. An unprecedented number of trials have been done, initiated, or are planned to test new immune modulators for cancer therapy using a variety of modified response criteria. This guideline will allow consistent conduct, interpretation, and analysis of trials of immunotherapies.

Lancet Oncol 2017; 18: e143–52

This online publication has been corrected. The corrected version first appeared at thelancet.com/oncology on April 30, 2019

Canadian Cancer Trials Group, Queen's University, Kingston, ON, Canada (Prof L Seymour MD, Prof J Dancey MD); **EORTC Headquarters, Brussels, Belgium** (J Bogaerts PhD, S Litière PhD); **Translational Medicine, Merck & Co, Kenilworth, NJ, USA** (A Perrone MD); **Clinical Trials Imaging Consulting, LLC, Belle Mead, NJ, USA** (R Ford MD); **Department of Radiology, Columbia University Medical Center, New York, NY, USA** (Prof L H Schwartz MD); **New York Presbyterian Hospital, New York, NY, USA** (Prof L H Schwartz); **Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, USA** (Prof S Mandrekar PhD); **Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA, USA** (N U Lin MD, Prof F S Hodi MD); **Early Clinical Trials Development Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute, Bethesda, MD, USA** (A Chen MD); **Institut de Recherche International Servier, Paris, France** (P Therasse MD); **Department of Radiology and Nuclear Medicine, VU University Medical Center, Amsterdam, Netherlands** (Prof O S Hoekstra MD); **Diagnostic Imaging Branch, National Cancer Institute, Bethesda, MD, USA** (L K Shankar MD); **Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA** (J D Wolchok MD); **Weill Cornell Medical and Graduate Colleges, New York, NY, USA** (M Ballinger PhD); **Ludwig**

Introduction

Changes in tumour burden (termed response) are often used as surrogates of survival or quality of life;[1] consequently, validated and consistent criteria for defining response to treatment are crucial. In 2000, the Response Evaluation Criteria in Solid Tumours (RECIST) working group simplified the 1981 WHO response criteria[2] after validation in a large data warehouse.[3] In 2009, RECIST was refined to RECIST version 1.1.[4] The RECIST working group ensures that RECIST undergoes continuous testing, validation, and updates.[5–7]

Immune modulators are one of the most important classes of new anticancer therapeutics.[8–10] Cytotoxic T-lymphocyte antigen-4 (CTLA-4), programmed death-1 (PD-1), and programmed death ligand-1 (PD-L1) pathways are the most intensively studied,[11–17] and drugs that are active in these pathways have, since 2011, received marketing authorisation (for some drugs the authorisation is conditional, pending the completion of other studies) for melanoma, lung, bladder, renal, and head and neck cancer.[18–23] The novel mechanism of action of these drugs, with immune and T-cell activation, is postulated to lead to unusual patterns of response that resemble tumour flare but are more pronounced and more frequent than previously described responses. In early trials of immune-based therapeutics in melanoma, investigators described unique response patterns, termed pseudoprogression. Some patients whose disease met the criteria for disease progression based on traditional response criteria such as RECIST (an increase in the sum of measures of target lesions, unequivocal increase in non-target disease, or the appearance of new lesions) were noted to have late but deep and durable responses.[24–28] In 2009, modified response criteria based on WHO criteria (which include the collection of bidimensional measurements of target lesions) were proposed—the immune-related response criteria (irRC).[29] The major modification involved the inclusion of the measurements of new target lesions (each must be at least 5×5 mm in size; with a maximum of ten visceral lesions in total, up to five new lesions per organ, and five new cutaneous lesions) into disease assessments. In 2013, researchers published revised irRC using unidimensional measurements based on the original RECIST.[30] Subsequent recommendations, some published in abstract form, seem to incorporate RECIST 1.1 recommendations.[31–33] These recommendations are often referred to as irRECIST, but have not always been consistently applied, leading to concerns about the comparability of data and results across trials, difficulty with pooling databases, and poor clarity regarding whether new lesions were measured, and if so, how many were captured, and whether measures were incorporated into tumour burden. Recent trials (since 2010) have generally used RECIST-based immune criteria to assess responses to immunotherapies.

Because of the need to standardise and validate response criteria, the RECIST working group prospectively planned to create a warehouse of data from trials of immunotherapeutics to test and validate RECIST 1.1 and suggest modifications if required. During the planning and initial collection of the immunotherapeutic warehouse, it was apparent that most trials testing these drugs have typically used RECIST 1.1 to define the primary and secondary efficacy-based endpoints, and reserved irRC or their modified definition of RECIST for exploratory endpoints.[31,32] Additionally, substantial variability in which criteria were used was seen across clinical trials within pharmaceutical companies and cooperative groups, leading to serious concerns about interpretation of pooled datasets. Finally, most trials that used immune-modified criteria used

Institute for Cancer Research, New York, NY, USA (M Ballinger)**; Genentech Inc, San Francisco, CA, USA** (M Ballinger)**; Department of Radiology, Gustav Roussy Cancer Campus, Villejuif, France** (C Caramella MD)**; and Department of Medical Oncology, University Medical Center Groningen, Groningen, Netherlands** (Prof E G E de Vries MD)

Correspondence to: Prof Lesley Seymour, Canadian Cancer Trials Group, Queen's University, Kingston, ON K7L 3N6, Canada **lseymour@ctg.queensu.ca**

For more on the **RECIST working group** see http://www.eortc.org/recist/

See Online for appendix

independent imaging review by a commercial entity for those criteria, rather than investigator assessments. We think that response criteria should be applicable across all cancer clinical trials, including those done in the academic sector, where costly independent review is not feasible.

On the basis of these observations, the RECIST working group decided to develop a guideline for the use of a modified RECIST to ensure consistent design and data collection that would facilitate the ongoing collection of clinical trial data and ultimate validation, if indicated, of a modified RECIST 1.1 for immune-based therapeutics (termed iRECIST). These guidelines are not intended to define or guide clinical practice or treatment decisions, but rather to provide a consistent framework for the management of data collected in clinical trials of immune-based therapies. Treatment decisions rest with the patient and their health-care team.

Terminology

iRECIST is based on RECIST 1.1. Responses assigned using iRECIST have a prefix of "i" (ie, immune)—eg, "immune" complete response (iCR) or partial response (iPR), and unconfirmed progressive disease (iUPD) or confirmed progressive disease (iCPD) to differentiate them from responses assigned using RECIST 1.1. Similar nomenclature is used for stable disease (iSD). New lesions are assessed and subcategorised into those that qualify as target lesions (new lesion, target) or non-target lesions (new lesion, non-target).

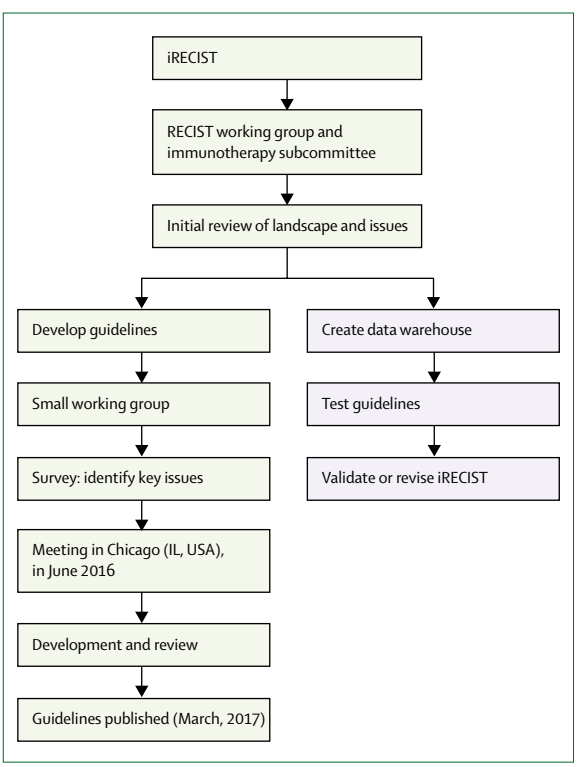


Figure 1: **Process for developing and validating iRECIST consensus guidelines**
Blue shaded boxes represent steps still in progress. RECIST=Response Evaluation Criteria in Solid Tumours.

Development of the guideline

The RECIST working group formed a subcommittee and held a series of conference calls and face-to-face meetings in 2015 and 2016 to discuss plans for the development and validation of iRECIST (figure 1) and to review existing approaches to assess response in immune modulator trials, and also to identify points of consensus and items that needed further discussion. Members of the subcommittee included clinical, statistical, and imaging experts in methodology and immunotherapy, representatives from the pharmaceutical companies developing immunotherapeutics, and key regulatory authorities (appendix p 1). On June 2, 2016, a formal meeting was held in Chicago (IL, USA), with invited presentations from regulatory authorities, pharmaceutical companies with immune modulator drugs in development, and academic groups, followed by a structured discussion. Before the meeting, the 52 invited participants were polled to enable the identification of questions that needed to be addressed, as well as the response criteria routinely used by participants. Ten respondents provided responses before the meeting (including some pooled responses) and all eight presenters identified additional areas of interest in their presentations. After review and discussion during the meeting, the group identified a list of important questions to be addressed by iRECIST (panel 1). Notably, all participants confirmed that RECIST 1.1 was used for primary endpoints, with immune-modified response criteria being used in an exploratory manner, with very few exceptions; in one instance, immune-modified criteria were used as a coprimary endpoint. The most commonly used immune-modified criteria were variations of irRECIST. There was more variability in independent imaging review and the period of time during which response data were collected after RECIST 1.1 progression or cessation of protocol therapy. Further calls and meetings were held to develop and plan the full validation of iRECIST (figure 1).

iRECIST

The continued use of RECIST 1.1 is recommended to define whether tumour lesions, including lymph nodes, are measurable or non-measurable, as well as for the management of bone lesions, cystic lesions, and lesions with previous local treatment (eg, radiotherapy; table 1). Similarly, no changes have been made to the recommendations regarding the method of measurement, although clinical examination and chest radiograph are rarely used, with the availability of more modern imaging techniques (eg, CT scans and MRI). The principles used to establish objective tumour response are largely unchanged from RECIST 1.1, but the major change for iRECIST is the concept of resetting the bar

if RECIST 1.1 progression is followed at the next assessment by tumour shrinkage.

iRECIST defines iUPD on the basis of RECIST 1.1 principles; however, iUPD requires confirmation, which is done on the basis of observing either a further increase in size (or in the number of new lesions) in the lesion category in which progression was first identified in (ie, target or non-target disease), or progression (defined by RECIST 1.1) in lesion categories that had not previously met RECIST 1.1 progression criteria. However, if progression is not confirmed, but instead tumour shrinkage occurs (compared with baseline), which meets the criteria of iCR, iPR, or iSD, then the bar is reset so that iUPD needs to occur again (compared with nadir values) and then be confirmed (by further growth) at the next assessment for iCPD to be assigned. If no change in tumour size or extent from iUPD occurs, then the timepoint response would again be iUPD. This approach allows atypical responses, such as delayed responses that occur after pseudoprogression, to be identified, further understood, and better characterised (tables 1–3, figure 2, appendix pp 2–4). Sample case record forms and protocol sections are included in the appendix pp 5–19. In the next few paragraphs, we only briefly summarise sections of RECIST 1.1 that are unchanged; readers should refer to RECIST 1.1 for full descriptions.[4]

Assessment of target, non-target, and new lesions

Most RECIST 1.1 recommendations are unchanged for timepoint response, including the management of lymph nodes, lesions that become too small to measure, lesions that split or coalesce, and the definition of complete response, partial response, stable disease, and progressive disease. Each timepoint response is based on the assessment of target lesions, non-target lesions, and new lesions.

For target lesions, iCR, iPR, and iSD can all be assigned after iUPD has been documented, as long as iCPD was not confirmed. iUPD is defined by RECIST 1.1 criteria for progressive disease; iUPD can be assigned multiple times as long as iCPD is not confirmed at the next assessment. Progression is confirmed in the target lesion category if the next imaging assessment after iUPD (4–8 weeks later) confirms a further increase in sum of measures of target disease from iUPD, with an increase of at least 5 mm. However, the criteria for iCPD (after iUPD) are not considered to have been met if complete response, partial response, or stable disease criteria (compared with baseline and as defined by RECIST 1.1) are met at the next assessment after iUPD. The status is reset (unlike RECIST 1.1, in which any progression precludes later complete response, partial response, or stable disease). iCR, iPR, or iSD should then be assigned; and if no change is detected, then the timepoint response is iUPD.

The assessment of non-target lesions at each timepoint follows similar principles. iUPD (but not iCPD) can have been documented before iCR or when the criteria for neither CR nor PD have been met (referred to as non-iCR/non-iUPD) and can be assigned several times, as long as iCPD was not confirmed. iUPD is defined by RECIST 1.1 criteria; however, iUPD can be assigned multiple times as long as iCPD is not confirmed at the next assessment. Progressive disease in the non-target lesion category is confirmed if subsequent imaging, done 4–8 weeks after iUPD, shows a further increase from iUPD. The criteria for iCPD are not judged to have

	RECIST 1.1	iRECIST
Definitions of measurable and non-measurable disease; numbers and site of target disease	Measurable lesions are ≥10 mm in diameter (≥15 mm for nodal lesions); maximum of five lesions (two per organ); all other disease is considered non-target (must be ≥10 mm in short axis for nodal disease)	No change from RECIST 1.1; however, new lesions are assessed as per RECIST 1.1 but are recorded separately on the case report form (but not included in the sum of lesions for target lesions identified at baseline)
Complete response, partial response, or stable disease	Cannot have met criteria for progression before complete response, partial response, or stable disease	Can have had iUPD (one or more instances), but not iCPD, before iCR, iPR, or iSD
Confirmation of complete response or partial response	Only required for non-randomised trials	As per RECIST 1.1
Confirmation of stable disease	Not required	As per RECIST 1.1
New lesions	Result in progression; recorded but not measured	Results in iUPD but iCPD is only assigned on the basis of this category if at next assessment additional new lesions appear or an increase in size of new lesions is seen (≥5 mm for sum of new lesion target or any increase in new lesion non-target); the appearance of new lesions when none have previously been recorded, can also confirm iCPD
Independent blinded review and central collection of scans	Recommended in some circumstances—eg, in some trials with progression-based endpoints planned for marketing approval	Collection of scans (but not independent review) recommended for all trials
Confirmation of progression	Not required (unless equivocal)	Required
Consideration of clinical status	Not included in assessment	Clinical stability is considered when deciding whether treatment is continued after iUPD

"i" indicates immune responses assigned using iRECIST. RECIST=Response Evaluation Criteria in Solid Tumours. iUPD=unconfirmed progression. iCPD=confirmed progression. iCR=complete response. iPR=partial response. iSD=stable disease.

Table 1: **Comparison of RECIST 1.1 and iRECIST**

been met if RECIST 1.1 criteria for complete response or non-iCR/non-iUPD are met after a previous iUPD. The status is reset (unlike RECIST 1.1) and iCR, or non-iCR/non-iUPD is assigned; if no change is detected, the timepoint response is iUPD.

RECIST 1.1 defines the appearance of new malignant lesions as denoting true disease progression, providing that other lesions (artefacts or benign intercurrent disease) are appropriately assessed and discounted if not malignant. These principles of RECIST 1.1 remain useful and clearly identify the management of new lesions that are considered to be potentially artefactual: "If a new lesion is equivocal, for example because of its small size, continued therapy and follow-up assessment will clarify whether it represents truly new disease. If repeat scans confirm there is definitely a new lesion, then progression should be declared using the date of the initial scan".[4]

However, many aspects of new lesion assessment are unique to iRECIST. If a new lesion is identified (thus meeting the criteria for iUPD) and the patient is clinically stable, treatment should be continued. New lesions should be assessed and categorised as measurable or non-measurable using RECIST 1.1 principles. Five lesions (no more than two per organ) should be measured and recorded as a new lesion target, but should not be included in the sum of measures of the original target lesions identified at baseline (appendix p 17). Other

***Panel 1:* Key questions identified by the RECIST working group**

- How to define the date of progression in scenarios in which initial progression by RECIST 1.1 is followed by response and later progression
- How to define best overall response when initial progression is established with RECIST 1.1
- How to manage response and progression in trials comparing standard non-immunotherapy drugs against immunotherapeutics
- Whether or not progression should be confirmed with a second scan; and if so, which timepoint denotes the date of progression?
- New lesions: when to measure, how many to measure, and whether all should be measured at each subsequent assessment
- Optimal timing of frequency of response assessment
- How to manage therapeutic interventions such as surgery or radiotherapy after response

	Timepoint response with no previous iUPD in any category	Timepoint response with previous iUPD in any category*
Target lesions: iCR; non-target lesions: iCR; new lesions: no	iCR	iCR
Target lesions: iCR; non-target lesions: non-iCR/non-iUPD; new lesions: no	iPR	iPR
Target lesions: iPR; non-target lesions: non-iCR/non-iUPD; new lesions: no	iPR	iPR
Target lesions: iSD; non-target lesions: non-iCR/non-iUPD; new lesions: no	iSD	iSD
Target lesions: iUPD with no change, or with a decrease from last timepoint; non-target lesions: iUPD with no change, or decrease from last timepoint; new lesions: yes	Not applicable	New lesions confirm iCPD if new lesions were previously identified and they have increased in size (≥5 mm in sum of measures for new lesion target or any increase for new lesion non-target) or number; if no change is seen in new lesions (size or number) from last timepoint, assignment remains iUPD
Target lesions: iSD, iPR, iCR; non-target lesions: iUPD; new lesions: no	iUPD	Remains iUPD unless iCPD is confirmed on the basis of a further increase in the size of non-target disease (does not need to meet RECIST 1.1 criteria for unequivocal progression)
Target lesions: iUPD; non-target lesions: non-iCR/non-iUPD, or iCR; new lesions: no	iUPD	Remains iUPD unless iCPD is confirmed on the basis of a further increase in sum of measures ≥5 mm; otherwise, assignment remains iUPD
Target lesions: iUPD; non-target lesions: iUPD; new lesions: no	iUPD	Remains iUPD unless iCPD is confirmed based on a further increase in previously identified target lesion iUPD in sum of measures ≥5 mm or non-target lesion iUPD (previous assessment need not have shown unequivocal progression)
Target lesions: iUPD; non-target lesions: iUPD; new lesions: yes	iUPD	Remains iUPD unless iCPD is confirmed on the basis of a further increase in previously identified target lesion iUPD sum of measures ≥5 mm, previously identified non-target lesion iUPD (does not need to be unequivocal), or an increase in the size or number of new lesions previously identified
Target lesions: non-iUPD or progression; non-target lesions: non-iUPD or progression; new lesions: yes	iUPD	Remains iUPD unless iCPD is confirmed on the basis of an increase in the size or number of new lesions previously identified

Target lesions, non-target lesions, and new lesions defined according to RECIST 1.1 principles; if no pseudoprogression occurs, RECIST 1.1 and iRECIST categories for complete response, partial response, and stable disease would be the same. *Previously identified in assessment immediately before this timepoint. "i" indicates immune responses assigned using iRECIST. iCR=complete response. iPR=partial response. iSD=stable disease. iUPD=unconfirmed progression. non-iCR/non-iUPD=criteria for neither CR nor PD have been met. iCPD=confirmed progression. RECIST=Response Evaluation Criteria in Solid Tumours.

***Table 2:* Assignment of timepoint response using iRECIST**

	Timepoint response 1	Timepoint response 2	Timepoint response 3	Timepoint response 4	Timepoint response 5	iBOR
Example 1	iCR	iCR, iUPD, or NE	iCR, iUPD, or NE	iUPD	iCPD	iCR
Example 2	iUPD	iPR, iSD, or NE	iCR	iCR, iUPD, or NE	iCR, iPR, iSD, iUPD, iCPD, or NE	iCR
Example 3	iUPD	iPR	iPR, iSD, iUPD, or NE	iPR, iSD, iUPD, NE, or iCPD	iPR, iSD, iUPD, NE, or iCPD	iPR
Example 4	iUPD	iSD or NE	iPR	iPR, iSD, iUPD, or NE	iPR, iSD, iUPD, iCPD, or NE	iPR
Example 5	iUPD	iSD	iSD, iUPD, or NE	iSD, iUPD, iCPD, or NE	iSD, iUPD, iCPD, or NE	iSD
Example 6	iUPD	iCPD	Any	Any	Any	iCPD
Example 7	iUPD	iUPD (no iCPD)	iCPD	Any	Any	iCPD
Example 8	iUPD	NE	NE	NE	NE	iUPD

Eight examples are presented for patients with target disease at baseline, but many more scenarios exist following the same principles. Table assumes a randomised study in which confirmation of complete response or partial response is not required. For patients with non-target disease only at baseline, only iCR or non-complete response or non-progression of disease can be assigned at each timepoint (not shown in the table for ease of presentation). "i" indicates immune responses assigned using iRECIST. iBOR=best overall response. iCR=complete response. iPR=partial response. NE=not evaluable. iUPD=unconfirmed progression. iCPD=confirmed progression. iSD=stable disease. RECIST=Response Evaluation Criteria in Solid Tumours.

Table 3: **Scenarios of assignments of best overall response using iRECIST**

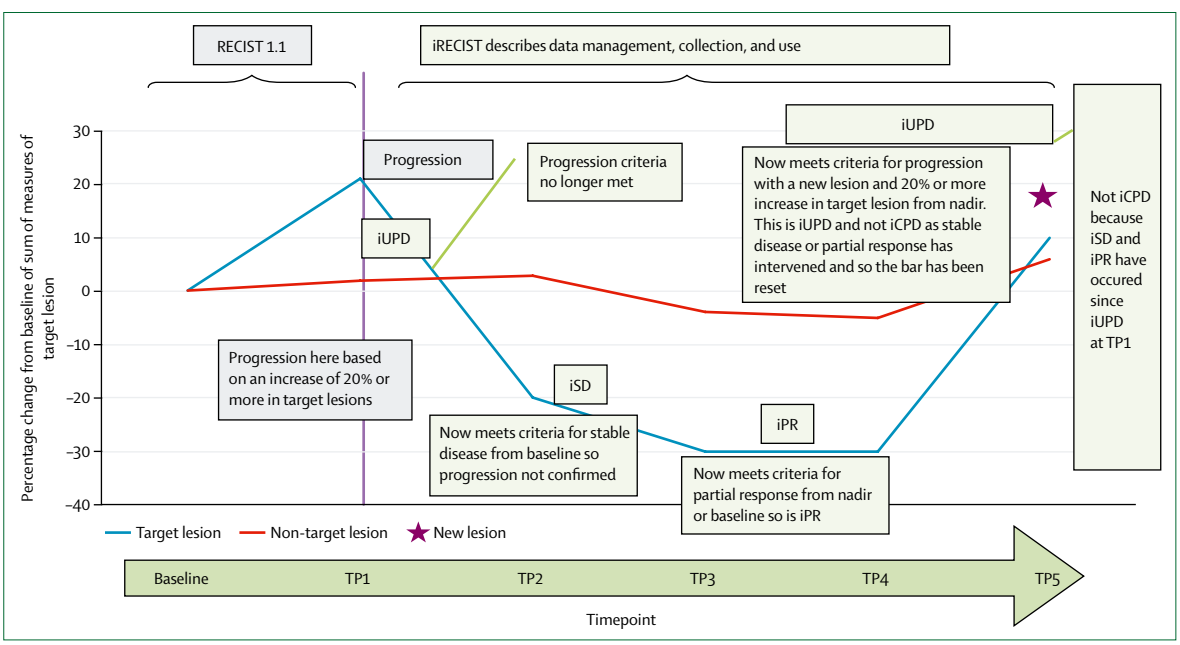


Figure 2: **RECIST 1.1 and iRECIST: an example of assessment**
Prefix "i" indicates immune responses assigned using iRECIST; others without "i" are confirmed by RECIST 1.1. RECIST=Response Evaluation Criteria in Solid Tumours. iCR=complete response. iCPD=complete progression. iPR=partial response. iSD=stable disease. iUPD=unconfirmed progression. TP=timepoint.

measurable and non-measurable lesions are recorded as new lesion non-target. Trialists might choose to measure and record more than five new lesions for research purposes, but this method is not believed to be practical for general use. New lesions do not need to meet the criteria for new lesion target to result in iUPD (or iCPD); new lesion non-target can also drive iUPD or iCPD. Progressive disease is confirmed (iCPD) in the new lesion category if the next imaging assessment, done at 4–8 weeks after iUPD, confirms additional new lesions or a further increase in new lesion size from iUPD (sum of measures increase in new lesion target ≥5 mm, any increase for new lesion non-target).

Notably, if iUPD criteria were met on the basis of progression in the target or non-target disease, or the appearance of new lesions, then RECIST 1.1-assigned progression in another lesion category in the confirmatory scan also confirms iCPD.

Continued treatment after iUPD

The existing literature describes pseudoprogression as an increase in the size of lesions, or the visualisation of new lesions, followed by a response, which might be durable. Although well described, differentiating transient pseudoprogression from true progression, potentially requiring a change in therapy, can be challenging. Although early discontinuation of an effective drug is not desirable, continued long-term treatment with a non-effective drug past true progression might delay the initiation of potentially effective salvage therapy.

***Panel 2:* Key principles to be considered**

- If the criteria for iUPD have never been met, principles follow RECIST 1.1
- However, if the criteria for iUPD have been met, the next timepoint response could be:
  - iUPD: no change noted in any category of lesion
  - iSD, iPR, or iCR. Here, iUPD (followed by iCPD) should occur again
  - iCPD, if the category in which iUPD was met at the last timepoint response shows a further increase in tumour burden as evidenced (as applicable) by a ≥5 mm increase in sum of measures of target or new target lesions, further increase in non-target or new non-target lesions, or an increase in the number of new lesions

iCPD of a category which did not meet criteria for iUPD now meets the criteria for RECIST 1.1 progression Prefix "i" indicates immune responses assigned using iRECIST. RECIST=Response Evaluation Criteria in Solid Tumours. iCR=complete response. iCPD=complete progression. iPR=partial response. iSD=stable disease. iUPD=unconfirmed progression.

We recommend that clinical trials in which treatment beyond initial RECIST 1.1-defined progression (ie, iUPD) is permitted should only allow patients who are clinically stable to continue on treatment until the next assessment (≥4 weeks later); this next imaging assessment should be no longer than 8 weeks later, to ensure that patients remain fit for salvage therapies. A longer timeframe before the next assessment might be reasonable if pseudoprogression is well described in the tumour type (eg, melanoma treated with a CTLA4 inhibitor), especially if no effective salvage therapies are available (eg, *BRAF* wild-type melanoma) but should be justified in the trial protocol. All decisions regarding continuation or discontinuation of therapy should be made by the patient and their health-care provider; iRECIST describes what data are to be collected, submitted, and analysed in clinical trials of immune-based therapies.

An assignment of clinical stability requires that no worsening of performance status has occurred, that no clinically relevant increases in disease-related symptoms such as pain or dyspnoea occur that are thought to be associated with disease progression (these symptoms are generally understood to mean a requirement for increased palliative intervention), and that no requirement for intensified management of disease-related symptoms exists, including increased analgesia, radiotherapy, or other palliative care.

The imaging findings and the recommendation to continue with treatment despite iUPD should be discussed with the patient before a decision is made about whether or not to continue therapy. Patients who have iUPD and are not clinically stable should be designated as not clinically stable in the case report form. This designation will allow the best overall response to be calculated and the date of iUPD to be used in estimates of progression-free survival.

If the confirmatory scan confirms iCPD, but the investigator or patient believes that continued treatment is appropriate, imaging should continue and data should be collected to allow further elucidation of tumour growth dynamics with immune modulators. For the same reason, and if feasible, even patients who discontinue therapy for iCPD are recommended to continue to have disease assessments until they start other systemic or local therapies.

Timepoint and best overall response

Although the principles of the assignment of the timepoint response and best overall response closely follow RECIST 1.1, and reflect assessment of target and non-target lesions as well as the presence of new lesions, the possibility of pseudoprogression adds complexity (tables 1–3, panel 2, appendix pp 2–4). The timepoint response is calculated using the response assigned for each category of lesion (as for RECIST 1.1), but takes into account the last timepoint response.

The algorithm for patients with no previous iUPD is identical to RECIST 1.1. For patients with iUPD at the last timepoint response, the next timepoint response is dependent on the status of all lesions, including target, non-target, new lesion target, and new lesion non-target; on whether any increase in size has occurred (either a further increase in size or a sufficient increase to assign a new iUPD if the criteria were not previously met); or the appearance of additional new lesions.

For iRECIST, the best overall response (iBOR) is the best timepoint response recorded from the start of the study treatment until the end of treatment, taking into account any requirement for confirmation. iUPD will not override a subsequent best overall response of iSD, iPR, or iCR (tables 1–3, appendix pp 2–4), meaning that iPR or iSD can be assigned (timepoint response or iBOR) even if new lesions have not regressed, or if unequivocal progression (non-target lesions) remains unchanged, providing that the criteria for iCPD are not met.

Confirmation of response is not required when using RECIST 1.1, except in non-randomised trials, and this approach is also recommended for iRECIST. The duration of iCR and iPR is from the timepoint when the criteria for iCR or iPR are first met, whereas the duration of iSD is still calculated from baseline.

The protocol should establish how missing response assessments will be handled. Assessments that are not done or are not evaluable should be disregarded. For example, an iUPD followed by an assessment that was not done or not evaluable, and then another unconfirmed progressive disease, would be indicative of iCPD. Protocols should clearly specify whether assessments done after protocol therapy is discontinued can be considered in identification of iBOR; it might be

reasonable to include assessments done several weeks or months after protocol treatment has been discontinued if late responses are anticipated (such as with a CTLA4 inhibitor) and patients have not received other systemic or local therapies. Protocols should also specify how any new therapy introduced before progression (eg, radiotherapy or surgery) will affect iBOR designation. Other RECIST 1.1 recommendations, including the management of missing assessments, remain unchanged, including requiring that the statistical analysis plan should indicate how missing data or assessments will be addressed in the determination of response and progression.

Frequency of tumour reassessment

In general, follow-up response assessment every 6–12 weeks is recommended for iRECIST, depending on the frequency of treatment visits, as recommended for RECIST 1.1. The protocol should specify which anatomical locations are assessed at baseline and follow-up, and whether bone scans should be repeated at each response assessment, only to confirm iPR or iCR, or when clinically indicated. For all trials, especially comparative ones, response assessments should be done on a calendar schedule and not be affected by delays in therapy or the requirement for earlier confirmatory scans, which might be done to confirm iUPD or in some trials, to confirm complete or partial response.

Tumour reassessment can be done earlier than originally planned (but only between 4 and 8 weeks after iUPD) to confirm iUPD (or, in non-randomised trials, to confirm iCR or iPR ≥4 weeks after the scan showing complete or partial response). If progression is not confirmed, reassessment should continue as originally planned (ie, if scans were to be done at 8, 16, and 24 weeks, and a scan was done at 12 weeks to confirm response, then the next scans should be done at 16 weeks and 24 weeks, as planned). If patients continue on treatment per protocol after iCPD, assessments should continue to be done, at the same planned schedule, until protocol treatment is discontinued.

Ideally, all imaging done after protocol treatment has been discontinued should continue to be recorded on the case report form until subsequent therapies are initiated, as the protocol and informed consent document permit. These data will allow further refinement of iRECIST.

Statistical and protocol considerations

The event date to be used for calculation of progression-free survival (iPFS) should be the first date at which progression criteria are met (ie, the date of iUPD) provided that iCPD is confirmed at the next assessment (appendix pp 2–4 and 19). If iUPD occurs, but is disregarded because of later iSD, iPR, or iCR, that iUPD date should not be used as the progression event date.

If progression is not confirmed and there is no subsequent iSD, iPR, or iCR, then the iUPD date should still be used in the following scenarios: if the patient stops protocol treatment because they were not judged to be clinically stable, or no further response assessments are done (because of patient refusal, protocol non-compliance, or patient death); the next timepoint responses are all iUPD, and iCPD never occurs; or the patient dies from their cancer. The case report form collects the reason why confirmatory response assessment was not done at any timepoint, such as not clinically stable, centre error, patient refusal, or patient death.

For protocols that permit crossover, or if intermittent schedules are being tested, the protocol should clearly specify whether iUPD or iCPD would be used for a treatment decision leading to crossover and how data subsequent to crossover will be managed and analysed. In general, we suggest that iCPD be used especially for scenarios with immunotherapy in both treatment groups and when pseudoprogression is anticipated.

Adjuvant trials of immune modulators given after curative surgery for melanoma or lung cancer are ongoing (NCT 02437279, 02388906, 02595944, 02504372, and 02273375) but have yet to report their results. Suspected new lesions in the curative setting should always be investigated thoroughly and preferably have a biopsy taken before the designation of relapse is assigned. If taking a biopsy sample is not technically feasible, then it would seem to be reasonable to follow the principles of iRECIST, with a follow-up scan to confirm relapse in patients who are clinically stable.

The collection of anonymised imaging (even if centralised blinded review of imaging studies is not planned) is recommended for all studies using an imaging-based endpoint (ie, response or progression-free survival) if feasible. Although the iRECIST guideline requires the recording of the measurements of up to five new lesions, it might eventually be necessary to record additional lesions to obtain a more precise estimate of progression. Central collection of images will allow further assessment by an independent radiologist if necessary. If real-time central review is planned, the protocol should clearly explain how treatment decisions will be made.

We recommend that phase 3 clinical trials continue to incorporate both RECIST 1.1 and iRECIST (table 1) and that RECIST 1.1 should continue to be used to define the primary efficacy outcomes (progression-free survival, disease progression, and best overall response). Exploratory analyses using the iPD date (ie, the first date of iUPD that is subsequently confirmed) can be defined in the statistical analysis plan. Early-phase trials can consider using iRECIST as the primary criteria. The protocol should carefully explain which will be the primary criteria used to assess response, and which would be exploratory. This information is especially important for trials that compare an immune modulator treatment with a non-immune modulator treatment.

Discussion: next steps and validation

Immunotherapeutics are a major advance in the treatment of an escalating number of cancers. The increasing testing and use of these drugs in multiple clinical settings, including adjuvant, first, second, and subsequent lines of therapy will require the use of progression-based endpoints. RECIST 1.1 might not always adequately capture the unique patterns of response that have been well described in clinical trials of these drugs in a low proportion of patients, typically reported as 10% or less, mainly in melanoma studies.[32–34] The true frequency in trials of other malignancies (including non-small-cell lung cancer) is unclear because most trials have reported RECIST 1.1-based response rates,[35] but might be less common based on anecdotal reports. Similarly, whether this pattern is unique to drugs active in the CTLA4–PD-1–PD-L1 pathway is currently unknown. Trials testing immunotherapeutics in combination with standard therapies, especially when they are compared with these standard therapies alone, further confound the assessment of progression-based endpoints.

RECIST 1.1 already addresses the management of equivocal progression, including suspected new lesions, which might explain, at least in part, the continued use of RECIST 1.1 to define response-based primary endpoints. RECIST 1.1 deals with mainly technical differences in scans that give the appearance that new lesions might have developed, or the concept of the isodense lesion at baseline that becomes more visible after the start of therapy since it becomes internally more necrotic as opposed to a true new lesion. However, the intention was never to use those recommendations to manage pseudoprogression described with immune modulators.

Although modified response criteria have been used, a formal guideline is clearly needed, with robust plans for prospective testing and consistent data collection and validation. Trials have not always been consistent in the definition of the response criteria to be used, have used trial-specific modifications of response criteria in which new lesion measurements can or cannot be included in the assessment of response, and response assessments after progression defined by RECIST 1·1 are not always done. Those data are crucial to understand the dynamics of tumour response to immunotherapeutics, including whether immunotherapeutics with different mechanisms of action have varying effects.

Although some progress has been made in understanding tumour dynamics with immunotherapeutics, progress in this area has undoubtedly been limited by reluctancy toward data sharing across trials, companies, and immunotherapeutics. Publications have been based on trials done by individual pharmaceutical companies or commercial organisations. In the development of this guideline, virtually all major pharmaceutical companies developing immunotherapeutics participated and have shared their experiences, protocols, response criteria, and, most importantly, their data. The iRECIST team also included members of the European Medicines Agency and the US Food and Drug Administration.

Although this guideline is consensus based, it is not yet validated because the data warehouse is still being created with initial trial data already in place. The guideline includes all available knowledge on response dynamics, allowing appropriate management of true pseudoprogression, but importantly, it also safeguards patients: although pseudoprogression is now well described, it still only occurs in fewer than one in ten patients. Treatment past radiographic progression might be appropriate only in a small number of patients, and the continuation of treatment past true progression could reduce subsequent effective therapies if the patient is no longer fit enough to tolerate any further treatment.

iRECIST requires the confirmation of progression to rule out or confirm pseudoprogression. Although this recommendation is in keeping with that of RECIST 1.1 to continue treatment and repeat imaging in the case of a mixed response or equivocal findings, if pseudoprogression is common, patients might be exposed to a higher risk (of continuing ineffective therapy or increasing exposure to radiotherapy) or cost (for the potentially ineffective therapy or the costs of imaging). We recommend that these criteria are used for clinical trial protocols rather than to guide clinical practice. Treatment beyond RECIST 1.1-based progression should be considered only in carefully selected scenarios in which the patient is stable (or improving) symptomatically and if there is just a short period remaining before reassessment.

Although at first glance the recommendation to collect measurements of new lesions as defined in this guideline seems onerous, the collection of these measurements and the recording of both RECIST 1.1 and iRECIST for timepoint response and best overall response have several benefits. The association between the site of the new lesion and progression-free survival and the value of adding new lesion measurements to the sum of measures can be explored. Continuing to record RECIST 1.1 allows comparison with reported immunotherapy trials that have used RECIST 1.1, as well as chemotherapy trials, while also allowing treatment past progression and collecting data that will allow further testing and validation of iRECIST. Differences in trial outcomes using RECIST 1.1 versus iRECIST could occur, and the interpretation will be informative. Our proposed plan will enable identification of such situations, and hopefully clarification of the underlying mechanisms. Additionally, in the future, quantification of the differences in outcome estimation between RECIST 1.1 and iRECIST will be possible, enabling better informed decisions for future changes to RECIST guidelines.

This strategy will also be useful for trials comparing immunotherapy-based with non-immunotherapy-based therapeutics. RECIST 1.1 and iRECIST should yield almost identical results for non-immunotherapy treatments,

Search strategy and selection criteria

This paper describes a consensus guideline, rather than a formal literature review. However, a database search was done using PubMed in August, 2016, with the following search terms: "immune response criteria" (limited to cancer, clinical trials, and publications in English language; 234 citations), "irRC" (23 citations), and "pseudoprogression" (limited to cancer, clinical trials, and pubications in English language; 39 citations).

based on the RECIST warehouses; whereas an immune modulator warehouse and associated sensitivity analysis of endpoints will enable the quantification of potential added benefit for the immunotherapy component. Although comparison of iRECIST in such situations incorporates an element of bias by construction, confirmation and validation of the guideline by overall survival results might gain additional importance.

Our recommendation for the design of randomised studies planned for licensing applications is to continue to use RECIST 1.1 as the primary criteria for response-based endpoints. iRECIST should be regarded as exploratory in such trials, although earlier phase trials might consider using primarily iRECIST.

The creation of a data warehouse is underway and updates are available from EORTC where the warehouse is held. Meanwhile the implementation of this guideline, and the continued sharing of anonymised, patient-level data will allow the formal validation of iRECIST, ensuring that response-based guidelines remain robust and enable the rapid and robust future development of new cancer therapeutics to improve treatments for patients.

Contributors

All authors contributed to the literature search and writing of the report.

Declaration of interests

LS reports grants from AstraZeneca and Merck, outside the submitted work. AP owns stocks in Merck & Co. RF reports personal fees from Amgen, Abbvie, Aptiv, Aragon, BMS, Bioclinica, Celldex, Celsion, Clovis, Covance, Biomedical systems, ACR Image Metrix, Exelixis, Genentech, Janssen, Kyowa, Loxo, ICON Medical Imaging, Eisai, EMD Serono, Imaging Endpoints, Mlrati, Celgene, Merck, Novartis, Novocure, Roche, Pfizer, Quintile, Tokai, ONO, Red Hill, Radiant Sage, Orbimed Advisors, Cinven, Virtualscopics, Sun Advanced Pharma Research Company, Median Therapeutics, Optimer Biotechnology Inc, Vascular Biogenics, Ignyta, Immunocellular CBT Pharmaceuticals, Tracon, DNAtrix, CytRx, Kolltan, and Samsung Bioepsis, outside the submitted work. LHS reports grants and consulting fees from Novartis, grants from Astellas, Eli Lilly, Merck, Pfizer, consulting fees from GSK, outside the submitted work. JD reports grants from Merck, AstraZeneca, Pfizer, Ottawa Hospital Research Institute, Novartis, outside the submitted work. FSH reports grants from Bristol-Myers Squibb, and personal fees from Merck, Novartis, Genentech, and EMD Serono, outside the submitted work. Additionally, FSH has a patent MICA Related Disorders with royalties paid, and a patent Tumor Antigens and Uses Thereof issued. JDW reports grants from Bristol Myers Squibb, Merck, and Genentech during the conduct of the study; consulting fees from Bristol Myers Squibb and Merck, and is on the scientific advisory board for Medimmune, outside the submitted work. MB reports personal fees from Genentech outside the submitted work, and owns stocks in Roche. CC reports personal fees from Roche, BMS, and AstraZeneca, outside the submitted work. EGEdV reports consulting fees from Synthon, Medivation, and Merck; and grants from Novartis, Amgen, Roche/Genentech, Servier, Chugai, Synthon, AstraZeneca, and Radius Health, outside the submitted work. JB, SM, NUL, SL, AC, PT, OSH, and LKS declare no competing interests.

Acknowledgments

AC and LKS are employees of the National Institutes of Health/National Cancer Institute. This publication was supported by the Canadian Cancer Society Research Institute (grant #021039), the EORTC Cancer Research Fund, and the National Cancer Institute (grant number 5U10-CA11488-45). The contents of this paper were presented in part at the EORTC–NCI–AACR 2016 Meeting (Munich, Germany; Nov 29–Dec 2, 2016). We gratefully acknowledge the thoughtful participation of the following in this initiative: Patricia Keegan (US Food and Drug Administration, Silver Spring, MD, USA); Francesco Pignatti (European Medicine Agency, London, UK); Wendy Hayes (Bristol-Myers Squibb, Princeton, NJ, USA); Eric Rubin (Merck & Co, Kenilworth, NJ, USA). We also received written comments from Darragh Halpenny, Jean-Yves Blay, Florian Lordick, Silke Gillessen, Hirokazu Watanabe, Jose Pablo Maroto Rey, Pietro Quaglino, Howard Kaufman, Denis Lacombe, Corneel Coens, Catherine Fortpied, Jessica Menis, Francisco Vera-Badillo, Jean Powers, Michail Ignatiadis, Eric Gauthier, Michael O'Neal, Caroline Malhaire, Laure Fournier, and Glen Laird. We thank Anouk Funke for her assistance with this manuscript.

References

1 Kim C, Prasad V. Cancer drugs approved on the basis of a surrogate end point and subsequent overall survival: an analysis of 5 years of US Food and Drug Administration approvals. *JAMA Intern Med* 2015; **175:** 1992–94.
2 Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981; **47:** 207–14.
3 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* 2000; **92:** 205–16.
4 Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1·1). *Eur J Cancer* 2009; **45:** 228–47.
5 Liu Y, Litière S, de Vries EG, et al. The role of response evaluation criteria in solid tumor in anticancer treatment evaluation: results of a survey in the oncology community. *Eur J Cancer* 2014; **50:** 260–66.
6 Schwartz LH, Seymour L, Litière S, et al. RECIST 1.1—standardisation and disease-specific adaptations: perspectives from the RECIST working group. *Eur J Cancer* 2016; **62:** 138–45.
7 Schwartz LH, Litière S, de Vries E, et al. RECIST 1·1—update and clarification: from the RECIST committee. *Eur J Cancer* 2016; **62:** 132–37.
8 Brahmer J, Reckamp KL, Baas P, Crinò L, et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N Engl J Med* 2015; **373:** 123–35.
9 Pardoll, Drew M. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* 2012; **12:** 252–64.
10 Sidaway P. Bladder cancer: atezolizumab effective against advanced-stage disease. *Nat Rev Urol* 2016; **13:** 238.
11 Holt GE, Podack ER, Raez LE. Immunotherapy as a strategy for the treatment of non-small cell lung cancer. *Therapy* 2011; **8:** 43–54.
12 Sharma P, Wagner K, Wolchok JD, Allison JP. Novel cancer immunotherapy agents with survival benefit: recent successes and next steps. *Nat Rev Cancer* 2011; **11:** 805–12.
13 Antonia S, Goldberg SB, Balmanoukian A, et al. Safety and antitumor activity of durvalumab plus tremelimumab in non-small cell lung cancer: a multicentre, phase 1b study. *Lancet Oncol* 2016; **3:** 299–308.
14 Postow MA, Chesney J, Pavlick AC, et al. Nivolumab and ipilimumab versus ipilimumab in untreated melanoma. *N Engl J Med* 2015; **372:** 2006–17.
15 Reck M, Bondarenko I, Luft A, et al. Ipilimumab in combination with paclitaxel and carboplatin as first-line therapy in extensive-disease-small-cell lung cancer: results from a randomized, double-blind, multicenter phase 2 trial. *Ann Oncol* 2013; **24:** 75–83.

For the **EORTC RECIST data warehouse** see www.eortc.org/RECIST

16 Lynch TJ, Bondarenko I, Luft A, et al. Ipilimumab in combination with paclitaxel and carboplatin as first-line treatment in stage IIIB/IV non-small cell lung cancer: results from a randomized, double-blind, multi-center phase II study. *J Clin Oncol* 2012; **30:** 2046–54.

17 Larkin J, Chiarion-Sileni V, Gonzalez R, et al. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. *N Engl J Med* 2015; **373:** 23–34.

18 Hodi FS, O'Day SJ, McDermott DF, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 2010; **363:** 711–23.

19 Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Eng J Med* 2012; **366:** 2443–54.

20 Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer. *N Engl J Med* 2015; **373:** 1627–39.

21 Garon EB, Rizvi N, Hui R, et al. Pembrolizumab for the treatment of non-small cell lung cancer. *N Engl J Med* 2015; **372:** 2018–28.

22 Rosenberg JE, Hoffman-Censits J, Powles T, et al. Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *Lancet* 2016; **387:** 1909–20.

23 Motzer RJ, Escudier B, McDermott DF, et al. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med* 2015; **373:** 1803–13.

24 Fife BT, Bluestone JA. Control of peripheral T-cell tolerance and autoimmunity via CTLA-4 and PD-1 pathways. *Immunol Rev* 2008; **224:** 166–82.

25 Fife BT, Pauken KE, Eagar TN, et al. Interactions between PD-1 and PD-L1 promote tolerance by blocking the TCR-induced stop signal. *Nat Immunol* 2009; **10:** 1185–92.

26 Iwai Y, Ishida M, Tanaka Y, et al. Involvement of PD-L1 on tumor cells in the escape from host immune system and tumor immunotherapy by PD-L1 blockade. *Proc Natl Acad Sci USA* 2002; **99:** 12293–97.

27 Tarhini AA. Tremelimumab: a review of development to date in solid tumors. *Immunotherapy* 2013; **5:** 215–29.

28 Hoos A, Parmiani G, Hege K, et al. A clinical development paradigm for cancer vaccines and related biologics. *J Immunother* 2007; **30:** 1–15.

29 Wolchok JD, Hoos A, O'Day S, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res* 2009; **15:** 7412–20.

30 Nishino M, Giobbie-Hurder A, Gargano M, Suda M, Ramaiya NH, Hodi FS. Developing a common language for tumor response to immunotherapy: immune-related response criteria using unidimensional measurements. *Clin Cancer Res* 2013; **19:** 3936–43.

31 Bohnsack O, Ludajic K, Hoos A. Adaptation of the immune-related response criteria: irRECIST. *Ann Oncol* 2014; **25** (suppl 4)**:** iv361–iv372.

32 Hodi FS, Hwu WJ, Kefford R, et al. Evaluation of immune-related response criteria and RECIST v1·1 in patients with advanced melanoma treated with pembrolizumab. *J Clin Oncol* 2016; **34:** 1510–17.

33 Chiou VL, Burotto M. Pseudoprogression and immune-related response in solid tumors. *J Clin Oncol* 2015; **33:** 3541–43.

34 Kurra V, Sullivan RJ, Gainor JF, et al. Pseudoprogression in cancer immunotherapy: rates, time course and patient outcomes. *Proc Am Soc Clin Oncol* 2016; **34:** abstr 6580.

35 Reck M, Rodríguez-Abreu D, Robinson AG, et al. Pembrolizumab versus chemotherapy for PD-L1–positive non–small-cell lung cancer. *N Engl J Med* 2016; **375:** 1823–33

JAMA Oncology | Original Investigation

Hyperprogressive Disease in Patients With Advanced Non–Small Cell Lung Cancer Treated With PD-1/PD-L1 Inhibitors or With Single-Agent Chemotherapy

Roberto Ferrara, MD; Laura Mezquita, MD, PhD; Matthieu Texier, MSc; Jihene Lahmar, MD; Clarisse Audigier-Valette, MD; Laurent Tessonnier, MD; Julien Mazieres, MD, PhD; Gerard Zalcman, MD, PhD; Solenn Brosseau, MD; Sylvestre Le Moulec, MD; Laura Leroy, MD; Boris Duchemann, MD; Corentin Lefebvre, MD; Remi Veillon, MD; Virginie Westeel, MD, PhD; Serge Koscielny, MSc; Stephane Champiat, MD; Charles Ferté, MD, PhD; David Planchard, MD, PhD; Jordi Remon, MD; Marie-Eve Boucher, MD; Anas Gazzah, MD; Julien Adam, MD, PhD; Emilio Bria, MD; Giampaolo Tortora, MD, PhD; Jean-Charles Soria, MD, PhD; Benjamin Besse, MD, PhD; Caroline Caramella, MD

IMPORTANCE Hyperprogressive disease (HPD) is a new pattern of progression recently described in patients with cancer treated with programmed cell death 1 (PD-1) and programmed cell death ligand 1 (PD-L1) inhibitors. The rate and outcome of HPD in advanced non–small cell lung cancer (NSCLC) are unknown.

OBJECTIVES To investigate whether HPD is observed in patients with advanced NSCLC treated with PD-1/PD-L1 inhibitors compared with single-agent chemotherapy and whether there is an association between treatment and HPD.

DESIGN, SETTING, AND PARTICIPANTS In this multicenter retrospective study that included patients treated between August 4, 2011, and April 5, 2017, the setting was pretreated patients with advanced NSCLC who received PD-1/PD-L1 inhibitors (8 institutions) or single-agent chemotherapy (4 institutions) in France. Measurable disease defined by Response Evaluation Criteria in Solid Tumors (RECIST version 1.1) on at least 2 computed tomographic scans before treatment and 1 computed tomographic scan during treatment was required.

INTERVENTIONS The tumor growth rate (TGR) before and during treatment and variation per month (ΔTGR) were calculated. Hyperprogressive disease was defined as disease progression at the first evaluation with ΔTGR exceeding 50%.

MAIN OUTCOMES AND MEASURES The primary end point was assessment of the HPD rate in patients treated with IO or chemotherapy.

RESULTS Among 406 eligible patients treated with PD-1/PD-L1 inhibitors (63.8% male), 46.3% (n = 188) were 65 years or older, 72.4% (n = 294) had nonsquamous histology, and 92.9% (n = 377) received a PD-1 inhibitor as monotherapy in second-line therapy or later. The median follow-up was 12.1 months (95% CI, 10.1-13.8 months), and the median overall survival (OS) was 13.4 months (95% CI, 10.2-17.0 months). Fifty-six patients (13.8%) were classified as having HPD. Pseudoprogression was observed in 4.7% (n = 19) of the population. Hyperprogressive disease was significantly associated with more than 2 metastatic sites before PD-1/PD-L1 inhibitors compared with non-HPD (62.5% [35 of 56] vs 42.6% [149 of 350]; *P* = .006). Patients experiencing HPD within the first 6 weeks of PD-1/PD-L1 inhibitor treatment had significantly lower OS compared with patients with progressive disease (median OS, 3.4 months [95% CI, 2.8-7.5 months] vs 6.2 months [95% CI, 5.3-7.9 months]; hazard ratio, 2.18 [95% CI, 1.29-3.69]; *P* = .003). Among 59 eligible patients treated with chemotherapy, 3 (5.1%) were classified as having HPD.

CONCLUSIONS AND RELEVANCE Our study suggests that HPD is more common with PD-1/PD-L1 inhibitors compared with chemotherapy in pretreated patients with NSCLC and is also associated with high metastatic burden and poor prognosis in patients treated with PD-1/PD-L1 inhibitors. Additional studies are needed to determine the molecular mechanisms involved in HPD.

JAMA Oncol. 2018;4(11):1543-1552. doi:10.1001/jamaoncol.2018.3676
Published online September 6, 2018.

Author Audio Interview

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Benjamin Besse, MD, PhD, Cancer Medicine Department, Gustave Roussy, 114 Rue Edouard Vaillant, 94805 Villejuif, France (benjamin.besse@gustaveroussy.fr).

In the era of immuno-oncology, programmed cell death 1 (PD-1) and programmed cell death ligand 1 (PD-L1) inhibitors have demonstrated a clear survival benefit as a single agent or in combination compared with standard chemotherapy in both treatment-naive patients[1-4] and patients previously treated[5-8] for advanced non–small cell lung cancer (NSCLC). However, progression rates reported with single-agent PD-1/PD-L1 inhibitors are in some cases equal to or higher than with conventional treatment, ranging from 33% to 44% in pretreated patients with NSCLC.[5-7] Recently, an acceleration of tumor growth during immunotherapy, defined as hyperprogressive disease (HPD), was reported in 9% of advanced cancers[9] and in 29% of patients with head and neck cancer[10] treated with PD-1/PD-L1 inhibitors.

The tumor growth rate (TGR) is a tool for estimating the increase in tumor volume over time based on 2 computed tomography (CT) scan measurements.[11] The TGR takes into account the sum of the target lesions defined by Response Evaluation Criteria in Solid Tumors (RECIST version 1.1) and the interval between 2 CT scans. It can be used to quantitatively assess tumor dynamics and kinetics during treatment; specifically, it can be applied to identify the subset of patients experiencing HPD.

To explore if HPD is an unforeseen pattern of progression during IO therapy in NSCLC, we compared the TGR before and during IO therapy in a cohort of pretreated patients with advanced NSCLC. To investigate if HPD is a specific PD-1/PD-L1 inhibitor pattern, we assessed the TGR and HPD prevalence among a control cohort receiving single-agent chemotherapy.

Key Points

Question Do programmed cell death 1 (PD-1) and programmed cell death ligand 1 (PD-L1) inhibitors accelerate tumor growth, a phenomenon defined as hyperprogressive disease?

Findings In this multicenter cohort study including 406 patients with advanced non–small cell lung cancer (NSCLC) treated with PD-1/PD-L1 inhibitors, hyperprogressive disease was observed in 13.8% (n = 56) of the population. Patients experiencing hyperprogression had significantly worse overall survival (3.4 months) compared with patients with progression not classified as hyperprogressive disease (6.2 months).

Meaning Hyperprogressive disease is a novel pattern of progression in patients receiving treatment with PD-1/PD-L1 inhibitors for NSCLC, of which patients and clinicians should be aware to properly select the best treatment and carefully monitor disease evolution.

Methods

Patients and Treatment

In this multicenter study, data were retrospectively collected from all consecutive eligible patients with advanced NSCLC treated with IO (nivolumab, pembrolizumab, atezolizumab, or durvalumab) from November 10, 2012, to April 5, 2017, in 8 French institutions. For the control cohort, equivalent data were collected in patients with advanced NSCLC failing a platinum-based regimen and treated with single-agent chemotherapy (taxanes, pemetrexed, vinorelbine tartrate, or gemcitabine chlorohydrate) from August 4, 2011, to June 13, 2016, in 4 French institutions.

To be eligible, patients had to be 18 years or older, with histologically or cytologically confirmed stage III or IV NSCLC and available CT scans for radiological evaluation. In the single-agent chemotherapy control cohort, patients who received previous treatment with IO were excluded. The PD-L1 expression was analyzed by immunohistochemistry on tumor cells in archived biopsy specimens, when available, and the cutoff for positivity was 1%. This study was approved by the institutional review board of Gustave Roussy, and informed consent from participants was not required because of the retrospective nature.

Radiological Evaluation

At least 2 CT scans before PD-1/PD-L1 inhibitor therapy or chemotherapy (baseline and the most recent scan before baseline) and 1 CT scan during treatment were mandatory for radiological evaluation. The baseline CT had to be performed within 6 weeks before initiating treatment, and a minimum of 2 weeks between CT scans was required. All CT scans were centrally reviewed by 2 senior radiologists (L.T. and C.C.). The target lesions were defined according to RECIST version 1.1. An extensive assessment of noneligibility for radiological evaluation was performed in 1 center (Gustave Roussy) to refine inclusion of patients in subsequent centers. Therefore, patients from other centers were included only if eligible for radiological evaluation (ie, availability of the required CT scans, adequate intervals between them, and the presence of the target lesions). In cases of progression, if the patient was clinically stable, PD-1/PD-L1 inhibitors could be continued, with a subsequent evaluation at least 4 weeks later, according to immunotherapy response criteria recommendations.[12] Pseudoprogression was defined as initial progression, followed by complete response or partial response or stable disease lasting more than 6 months.[13]

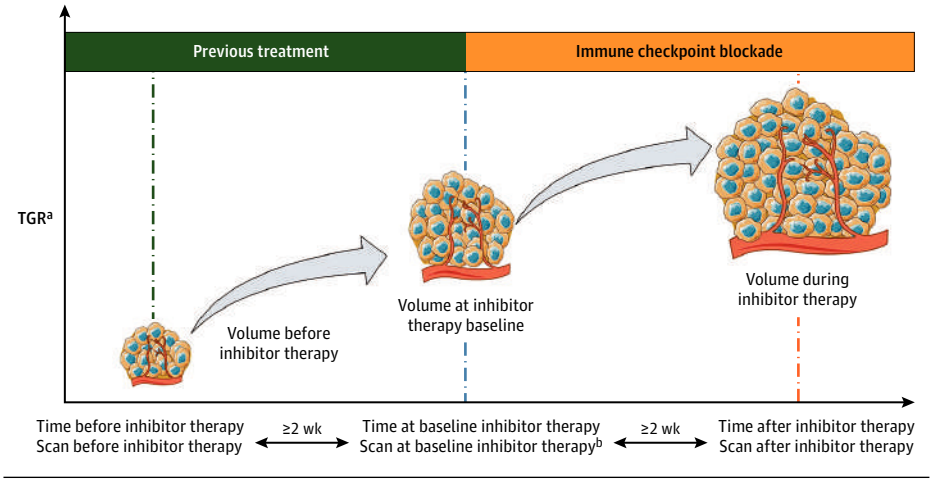
Tumor Growth Rate

The TGR was calculated according to the definition by Ferté et al[14] and was computed from the sum of the largest diameters of the target lesions as per RECIST version 1.1 (eMethods in the Supplement). The TGR results were reported as a percentage increase in tumor volume per month. New lesions and nonmeasurable disease were excluded from the RECIST version 1.1 sum, and the TGR was only quantified for the target lesions.[14]

The TGR was measured before and after PD-1/PD-L1 inhibitors (or chemotherapy in the control cohort). The difference (ΔTGR is the TGR on treatment minus the TGR before treatment) was used to assess the association of treatment with tumor growth. Delta TGR exceeding 0% means that treatment may accelerate tumor growth.

Hyperprogressive disease was defined as RECIST version 1.1 progressive disease on the first CT scan during treatment

Figure 1. Hypothetical Tumor Volume Variation and Definition of Hyperprogressive Disease (HPD) in the Immunotherapy Cohort



Variation of tumor growth rate (TGR) volume per month was calculated both before the start of programmed cell death (PD-1) and programmed cell death ligand 1 (PD-L1) inhibitor therapy and during PD-1/PD-L1 inhibitor therapy. Hyperprogressive disease was defined as Response Evaluation Criteria in Solid Tumors (RECIST) version 1.1 progressive disease at the first computed tomography (CT) scan during PD-1/PD-L1 inhibitor therapy and an absolute increase in the TGR exceeding 50% per month.
[a]New lesions and nonmeasurable disease not included in the TGR.
[b]The baseline CT scan performed within 6 weeks before PD-1/PD-L1 inhibitor therapy initiation.

and ΔTGR exceeding 50%, corresponding to an absolute increase in the TGR exceeding 50% per month. A graphical representation of the hypothetical tumor volume variation and the HPD definition for the immunotherapy cohort is shown in **Figure 1**.

Statistical Analysis

Associations between HPD and categorical or continuous variables were evaluated using the Fisher exact test and the *t* test, respectively. Because the diagnosis of HPD depends on the timing of the radiological assessment and could induce a lead-time bias,[15] a landmark analysis was performed to assess the association of HPD with overall survival (OS) using a time point at 6 weeks after PD-1/PD-L1 inhibitor or chemotherapy initiation. Patients alive at this time point and with progression on their first CT scan during PD-1/PD-L1 inhibitor therapy or chemotherapy were considered hyperprogressors or not hyperprogressors according to the diagnosis of HPD within the first 6 weeks of treatment. Overall survival curves were estimated with the Kaplan-Meier method and compared by the log-rank test. The hazard ratio (HR) was estimated using the univariate Cox proportional hazards regression model. All *P* values were 2 sided, and values less than .05 were considered statistically significant. Statistical analyses were performed using a software program (SAS for Windows, version 9.4; SAS Institute Inc).

Results

Immunotherapy Cohort

Overall, 406 patients (63.8% male) were included in the TGR analysis. The reasons for exclusion were evaluated in a single-center cohort (at Gustave Roussy, Villejuif, France) (n = 249) and included the following: unavailability of CT scans before baseline, at baseline, or during PD-1/PD-L1 inhibitor therapy; inadequate intervals between CT scans; or the absence of measurable disease. Of 249 patients, 76 (30.5%) were not evaluable for the TGR analysis, among whom 13.3% (33 of 249) experienced clinical progression and/or death before the first tumor evaluation during PD-1/PD-L1 inhibitor therapy (eFigure 1 in the Supplement).

The main characteristics of the 406 patients in the immunotherapy multicenter cohort are listed in the **Table**. The median follow-up was 12.1 months (95% CI, 10.1-13.8 months), the objective response rate was 18.9% (77 of 406), and 41.9% (170 of 406) of patients had progressive disease as the best response to immunotherapy (eTable 1 in the Supplement). The median progression-free survival (PFS) and OS were 2.1 months (95% CI, 1.8-3.1 months) and 13.4 months (95% CI, 10.2-17.0 months), respectively.

Before immunotherapy, 75 of 406 patients (18.5%) had a TGR of 0 or less (eTable 2 in the Supplement), but all were classified as having progressive disease because of the appearance of new lesions or progression in the nontarget lesions. During immunotherapy, the TGR was stable or decreased (ΔTGR ≤0) in 266 patients (65.5%) and increased (ΔTGR >0) in 140 patients (34.5%). Among them, 62 patients (15.3% of the overall population) were initially classified as having HPD (**Figure 2**A and **Figure 3**).

Overall, 19 patients (4.7%) had progressive disease, followed by complete response and/or partial response or stable disease longer than 6 months, and were thus classified as pseudoprogressors (eTable 1 in the Supplement). Six pseudoprogressors were initially classified as having HPD on the first CT scan. Excluding these 6 patients from the 62 patients with HPD, the definitive rate of HPD was 13.8% (56 patients). Hyperprogressive disease was significantly associated with more than 2 metastatic sites before PD-1/PD-L1 inhibitors compared with non-HPD (62.5% [35 of 56] vs 42.6% [149 of 350]; *P* = .006) (Table). No significant differences were observed according to the baseline tumor burden, the number of previous lines of therapy (eFigure 2 in the Supplement), or age (Table).

In the landmark survival analysis, patients experiencing HPD within the first 6 weeks of beginning PD-1/PD-L1 inhibitor therapy (n = 23) had significantly lower OS compared with

Table. Patient Characteristics and Association Between HPD Status and Clinical Categorical Variables for Immunotherapy-Treated Patients With NSCLC

Variable	No./Total No. (%)			Fisher Exact Test *P* Value
	Total (N = 406)	Non-HPD (n = 350)	HPD (n = 56)	
Age, y				
≥65	188 (46.3)	166 (47.4)	22 (39.3)	.31
<65	218 (53.7)	184 (52.6)	34 (60.7)	
Smoking history				
Current/former	371 (91.4)	319 (91.1)	52 (92.9)	>.99
None	35 (8.6)	31 (8.9)	4 (7.1)	
Smoking exposure, pack-years				
≤30	136 (33.5)	115/312 (36.9)	21/50 (42.0)	.53
>30	226 (55.7)	197/312 (63.1)	29/50 (58.0)	
Missing	44 (10.8)	38	6	
Histology				
Nonsquamous	294 (72.4)	252 (72.0)	42 (75.0)	.75
Squamous	112 (27.6)	98 (28.0)	14 (25.0)	
Stage[a]				
III	70 (17.2)	61 (17.4)	9 (16.1)	>.99
IV	336 (82.8)	289 (82.6)	47 (83.9)	
PD-L1 status[b]				
Negative	39 (9.6)	32/105 (30.5)	7/12 (58.3)	.10
Positive	78 (19.2)	73/105 (69.5)	5/12 (41.7)	
Missing	289 (71.2)	245	44	
Molecular status				
ALK rearrangement	4 (1.0)	3/233 (1.3)	1/36 (2.8)	.34
EGFR mutation	16 (3.9)	16/233 (6.9)	0	
KRAS mutation	87 (21.4)	74/233 (31.8)	13/36 (36.1)	
Wild type[c]	104 (25.6)	88/233 (37.8)	16/36 (44.4)	
Other alterations	58 (14.3)	52/233 (22.3)	6/36 (16.7)	
Missing	137 (33.7)	117	20	
No. of molecular alterations				
0-1	218 (53.7)	185/227 (81.5)	33/36 (91.7)	.16
≥2	45 (11.1)	42/227 (18.5)	3/36 (8.3)	
Missing	143 (35.2)	123	20	
Type of treatment before PD-1/PD-L1 inhibitor therapy				
Platinum-based chemotherapy	229 (56.4)	192 (54.9)	37 (66.1)	.61
Chemoradiotherapy	17 (4.2)	17 (4.9)	0	
Pemetrexed	17 (4.2)	15 (4.3)	2 (3.6)	
Taxanes	44 (10.8)	39 (11.1)	5 (8.9)	
Other chemotherapy	43 (10.6)	37 (10.6)	6 (10.7)	
Targeted therapy[d]	12 (3.0)	11 (3.1)	1 (1.8)	
Tyrosine kinase inhibitors[e]	37 (9.1)	33 (9.4)	4 (7.1)	
Immunotherapy	3 (0.7)	2 (0.6)	1 (1.8)	
No prior therapy	4 (1.0)	4 (1.1)	0	
Response to line before PD-1/PD-L1 inhibitor therapy				
Complete response/partial response	90 (22.2)	75/344 (21.8)	15/55 (27.3)	.08
Stable disease	185 (45.6)	167/344 (48.5)	18/55 (32.7)	
Progressive disease	124 (30.5)	102/344 (29.7)	22/55 (40.0)	
Missing	7 (1.7)	6	1	

(continued)

Table. Patient Characteristics and Association Between HPD Status and Clinical Categorical Variables for Immunotherapy-Treated Patients With NSCLC (continued)

Variable	No./Total No. (%)			Fisher Exact Test *P* Value
	Total (N = 406)	Non-HPD (n = 350)	HPD (n = 56)	
PD-1/PD-L1 inhibitor therapy line, range 1-9				
≤2[f]	218 (53.7)	186 (53.1)	32 (57.1)	.67
>2	188 (46.3)	164 (46.9)	24 (42.9)	
No. of metastatic sites before PD-1/PD-L1 inhibitor therapy				
≤2	222 (54.7)	201 (57.4)	21 (37.5)	.006
>2	184 (45.3)	149 (42.6)	35 (62.5)	
Type of inhibitor				
PD-1	377 (92.9)	325 (92.9)	52 (92.9)	>.99
PD-L1	29 (7.1)	25 (7.1)	4 (7.1)	
Monotherapy or combination				
Monotherapy	380 (93.6)	326 (92.9)	54 (96.4)	.56
Combination[g]	26 (6.4)	24 (6.9)	2 (3.6)	
ECOG performance status				
0-1	360 (88.7)	311 (88.9)	49 (87.5)	.82
≥2	46 (11.3)	39 (11.1)	7 (12.5)	
Subsequent therapy				
No	111 (27.3)	86/215 (40.0)	25/54 (46.3)	.44
Yes	158 (38.9)	129/215 (60.0)	29/54 (53.7)	
PD-1/PD-L1 inhibitor therapy ongoing or missing	137 (33.7)	135	2	
Neutrophil count, /μL				
≤7500	209 (51.5)	188/254 (74.0)	21/31 (67.7)	.52
>7500	76 (18.7)	66/254 (26.0)	10/31 (32.3)	
Missing	121 (29.8)	96	25	
Derived neutrophil-to-lymphocyte ratio				
≤3	195 (48.0)	174/254 (68.5)	21/31 (67.7)	>.99
>3	90 (22.2)	80/254 (31.5)	10/31 (32.3)	
Missing	121 (29.8)	96	25	
Lactate dehydrogenase level				
≤Upper limit of normal[h]	150 (36.9)	133/192 (69.3)	17/27 (63.0)	.51
>Upper limit of normal	69 (17.0)	59/192 (30.7)	10/27 (37.0)	
Missing	187 (46.1)	158	29	

Abbreviations: ECOG, Eastern Cooperative Oncology Group; EGFR, epidermal growth factor receptor; HPD, hyperprogressive disease; NSCLC, non–small cell lung cancer; PD-1, programmed cell death 1; PD-L1, programmed cell death ligand 1.

SI conversion factor: To convert neutrophil count to 10^9/L, multiply by 0.001.

[a] TNM stage (seventh edition) at advanced disease detection.

[b] Immunohistochemistry cutoff for positivity on tumor cells of 1% or higher.

[c] Wild type for *ALK* rearrangement, *EGFR* mutation, and *KRAS* mutation.

[d] In oncogene-addicted NSCLC.

[e] In non–oncogene-addicted NSCLC.

[f] Four patients treated in first line for metastatic disease.

[g] Combination of PD-1/PD-L1 inhibitor therapy and anti-EGFR monoclonal antibodies, PD-1/PD-L1 inhibitor therapy and chemotherapy, or PD-1/PD-L1 inhibitor therapy and immunotherapy (patients enrolled in clinical trials).

[h] Upper limit of normal defined according to the cutoff of each center.

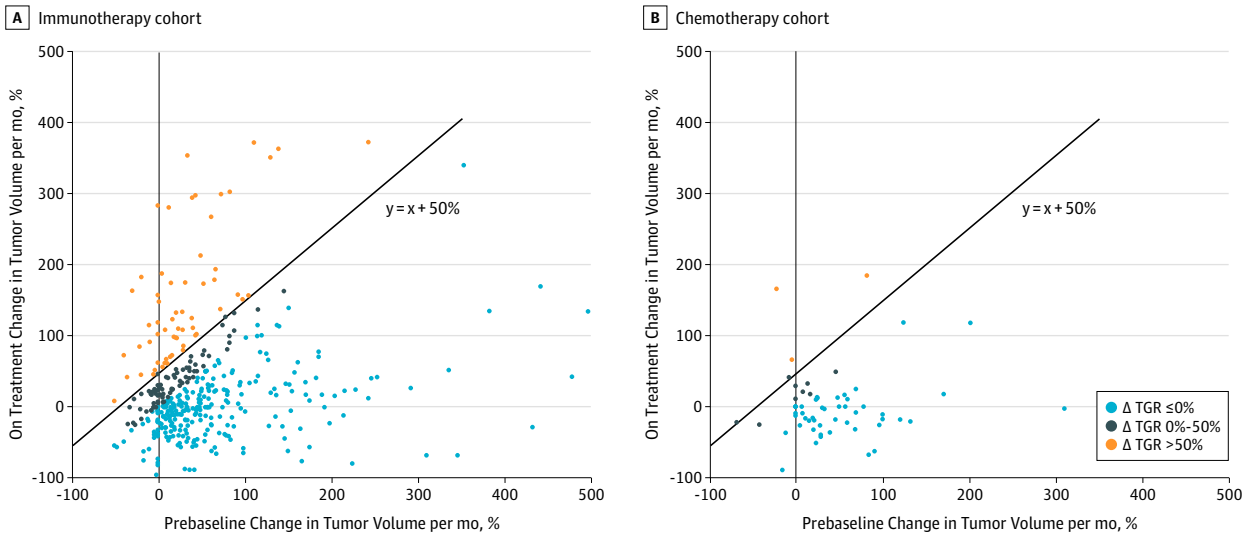
other patients with progressive disease (ie, non-HPD patients with progressive disease at the first evaluation [n = 138]) (median OS, 3.4 months [95% CI, 2.8-7.5 months] vs 6.2 months [95% CI, 5.3-7.9 months]; HR, 2.18 [95% CI, 1.29-3.69]; *P* = .003) (**Figure 4**A). As a sensitivity analysis, 2 other landmark time points were tested. With a time point at 4 weeks, the difference in OS remained significant. However, when choosing a time point of 8 weeks, the difference in OS did not reach statistical significance.

Chemotherapy Cohort

Overall, 59 patients were included in the TGR analysis. The reasons for exclusion were evaluated in a single-center cohort (at Gustave Roussy) (n = 77) (eFigure 3 in the Supplement). The main characteristics of the 59 patients are listed in eTable 3 in the Supplement. The median follow-up was 26.3 months (95% CI, 22.6-35.5 months), the objective response rate was 10.2% (6 of 59), and 30.5% (18 of 59) of patients had progressive disease as the best response (eTable 1 in the Supplement). The median PFS and OS were 3.9 months (95% CI, 3.1-4.8 months) and 8.6 months (95% CI, 6.2-13.4 months), respectively. No pseudoprogression was observed.

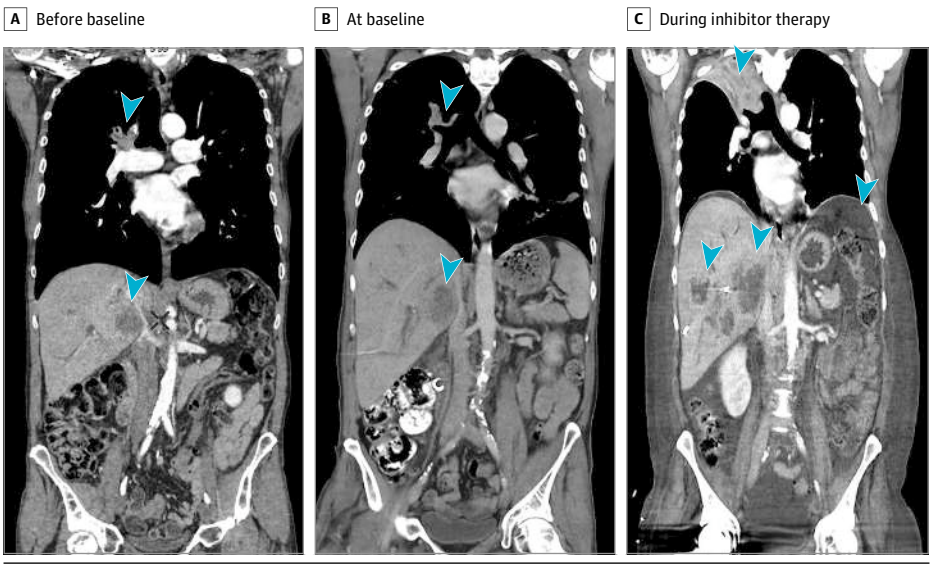
The TGR analysis is summarized in eTable 2 in the Supplement. Delta TGR was greater than 0 in 12 patients; among them, 3 patients were classified as having HPD (Figure 2B). A landmark analysis at 6 weeks showed a median OS of 4.5 months (95% CI, 2.5-6.5 months) in patients diagnosed as having HPD (n = 3) and 3.9 months (95% CI, 2.7-6.9 months) in other patients with progressive disease (ie, non-HPD patients with progressive disease at the first evaluation [n = 18]) (*P* = .60) (Figure 4B).

Figure 2. Scatterplots With Response According to Delta Tumor Growth Rate (TGR) in the Immunotherapy and Chemotherapy Cohorts



A, Light blue spots show 266 patients with regressing or stable tumors, dark blue spots show 78 patients with progressing tumors, and orange spots show 62 patients with accelerated tumor growth during programmed cell death (PD-1) and programmed cell death ligand 1 (PD-L1) inhibitor therapy. B, Light blue spots show 47 patients with regressing or stable tumors, dark blue spots show 9 patients with progressing tumors, and orange spots show 3 patients with accelerated tumor growth during chemotherapy. Diagonal lines separate patients with delta TGR exceeding 50% from patients with delta TGR of 50% or less.

Figure 3. Case Study of a Patient With Non–Small Cell Lung Cancer With Hyperprogressive Disease During Treatment With a PD-1 Inhibitor



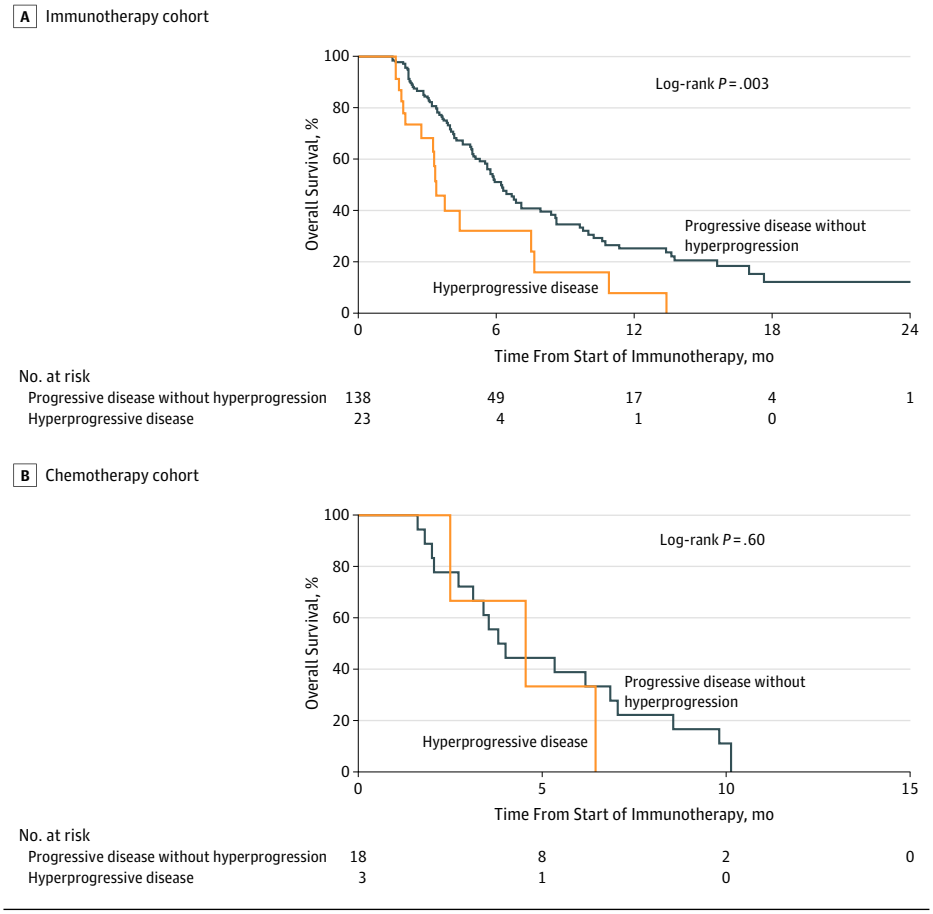
Shown are computed tomographic scans before baseline (A), at baseline about 3 weeks later (B), and during programmed cell death (PD-1) and programmed cell death ligand 1 (PD-L1) inhibitor therapy 1 month later (C) in a man in his mid-50s with stage IV (lung, liver, and bone metastases) *HER2*-amplified lung adenocarcinoma treated with anti–PD-1 therapy in the third line. After 2 administrations, there was evidence of extensive lung, liver, and peritoneal progression. Arrowheads show lung and liver metastases before and during anti–PD-1 treatment.

Discussion

In this study of pretreated patients with advanced NSCLC, HPD was observed in 13.8% (56 of 406) of patients treated with PD-1/PD-L1 inhibitors compared with 5.1% (3 of 59) of patients treated with single-agent chemotherapy. Our rate of HPD is concordant with the few relevant previously reported studies. Champiat et al[9] identified HPD in 9% of 131 patients with advanced cancer treated with PD-1/PD-L1 inhibitors in phase 1 trials; only 13 patients had lung cancer, and none were classified as having HPD. Hyperprogressive disease was identified more frequently (29%) by Saâda-Bouzid et al[10] among 34 patients with recurrent or metastatic head and neck cancer. This higher rate could have occurred because of the tumor type and/or their different definition of HPD.

In our study, HPD was associated with poor survival in patients with NSCLC treated with PD-1/PD-L1 inhibitors. Hyperprogressive disease could potentially explain the initial excess of death in some phase 3 trials. For example, in the CheckMate

Figure 4. Overall Survival for Hyperprogressive Disease (HPD) Compared With Progressive Disease Without Hyperprogression in the Immunotherapy and Chemotherapy Cohorts (6-Weeks Landmark Analysis)



A, Fourteen patients and 2 patients with pseudoprogression were excluded from the progressive disease without hyperprogression and HPD cohorts, respectively. The median overall survival is shown for 138 patients with progressive disease without hyperprogression at the first evaluation vs for 23 patients with diagnosed HPD within the first 6 weeks of treatment. B, The median overall survival is shown for 18 patients with progressive disease without hyperprogression at the first evaluation vs for 3 patients with diagnosed HPD within the first 6 weeks of treatment.

057 study,[5] the progression rate was 44% with nivolumab and 29% with docetaxel, with an excess of 14 deaths during the first 3 months in the nivolumab arm.[16] As a result, OS curves crossed at 6 months, with an initial survival benefit in favor of docetaxel. In addition, a recent retrospective study[17] reported that approximately 15% of early deaths were due to disease progression during the first 3 months of nivolumab treatment in patients with advanced NSCLC. The European Medicines Agency recently included an alert in the summary of product information for nivolumab regarding treatment of patients with NSCLC with poor prognostic features or aggressive disease.[18] In our study, the absence of a significant survival difference using a landmark analysis at 8 weeks is likely because of the small number of patients with HPD alive at that time point and eligible for the landmark analysis. This finding further suggests that HPD is a rapid phenomenon, which leads to early death mostly in the first 2 months of treatment.

There is no consensus on the optimal definition of HPD. Champiat et al[9] defined HPD as progressive disease at the first evaluation in addition to an increase of at least 2-fold in the TGR during PD-1/PD-L1 inhibitor therapy compared with the TGR before PD-1/PD-L1 inhibitors. Saâda-Bouzid et al[10] described HPD as an increase of at least 2-fold in tumor growth kinetics after PD-1/PD-L1 inhibitor initiation, which measured the variation of the sum of the largest diameters of the target lesions per unit of time during immunotherapy compared with tumor growth kinetic before PD-1/PD-L1 inhibitors. We used a stringent definition of HPD that requires a high-volume increase per month to classify a patient as a hyperprogressor. For example, a tumor with a 20% volume increase per month before immunotherapy had to have a 70% increase per month during immunotherapy to be labeled as HPD. Despite the differences in methods, the present analysis and the 2 previous studies[9,10] highlight the importance of quantifying tumor growth speed to discriminate between progression due to natural history of the disease (the tumor growth speed is already high before the start of the new treatment) and progression due to the potential intrinsic association of the experimental treatment (the tumor growth speed is lower before the start of the new treatment). Unfortunately, the TGR assessment cannot be validated in published randomized studies because the radiological evaluations before the baseline CT scan data were not captured.

In our immunotherapy cohort, HPD was significantly associated with a high number of metastatic sites before PD-1/PD-L1 inhibitors, whereas no association with the baseline tumor burden was found. However, the target lesions defined by RECIST version 1.1 do not always perfectly reflect the whole tumor burden, especially in patients with nonmeasurable dis-

ease (lung lymphangitis, bone metastases, and pleural or peritoneal effusions). Furthermore, high lactate dehydrogenase levels and a derived neutrophil to lymphocyte ratio exceeding 3 were recently shown to negatively influence the survival outcome of patients with NSCLC treated with PD-1/PD-L1 inhibitors.[19] In our analysis, no significant association was found between these biomarkers and HPD status; however, lactate dehydrogenase levels and neutrophil counts were not available for 46.1% (187 of 406) and 29.8% (121 of 406) of patients, respectively. Contrary to what was observed by Champiat et al,[9] no significant association between HPD and age was found in our study, probably because of the different methods used to assess HPD. Recently, Kato et al[20] identified *EGFR* mutations and *MDM2* amplification as possible molecular predictors of HPD. In our cohort, none of 16 patients with *EGFR*-mutated lung adenocarcinoma experienced HPD. Despite the association between *EGFR* mutations and decreased benefit from immunotherapy in patients with NSCLC,[21] the potential role of *EGFR* mutations in driving HPD remains unknown. The phenomenon of disease progression acceleration was previously described in oncogene-addicted NSCLC after interruption of targeted agents, such as RAF,[22] ALK,[23] and EGFR[24] inhibitors. In the present analysis, no significant association between HPD and the type of previous therapy was found, minimizing the risk of such an association.

In 6 of the 62 patients with HPD (9.7%), initial HPD was further reclassified as pseudoprogression, a feature described in 4.7% (19 of 406) of our total population, in line with a recently published study in the same setting.[25] Variable rates of pseudoprogression have been reported in patients with NSCLC (2%-19%),[26,27] melanoma (4%-7%),[28,29] and renal cell carcinoma (1%-15%)[30-32] on PD-1/PD-L1 blockade. However, a comparison of these numbers should be interpreted with caution in the absence of a common definition of pseudoprogression across the studies.[33] We identified HPD in only 3 of 59 patients (5.1%) treated with single-agent chemotherapy, and no pseudoprogression was described, suggesting that these patterns are new and specific to PD-1/PD-L1 inhibitors.

To our knowledge, the present study is the largest analysis exploring HPD to date and is the first conducted in a dedicated NSCLC population. In addition, we believe that this is the only study to include a control cohort of chemotherapy-treated patients with NSCLC and is thus able to assess the negative association with survivial of HPD compared with conventional disease progression during PD-1/PD-L1 inhibitor therapy. Although in some immunotherapy trials[5,6,8] the first CT scan was performed at week 9, the fact that HPD drives toward early death (mainly in the first 6 weeks of treatment) prompts discussion over an anticipated first radiological evaluation during PD-1/PD-L1 inhibitor therapy to properly identify hyperprogressors. Ultimately, because of the poor OS associated with HPD, an early switch to salvage chemotherapy in these patients should be considered.

Limitations

Our study has some limitations, mainly related to its retrospective nature. First, a potential underestimation of HPD may have occurred because 30.5% (76 of 249) of the patients treated in 1 center (Gustave Roussy) were excluded from the TGR analysis, mostly because of rapid progression and/or death that prevented any further evaluation or because of the absence of the target lesions. In addition, in our study, PD-L1 expression was not available for 71.2% (289 of 406) of patients because this information was not mandatory for PD-1/PD-L1 inhibitor prescription in pretreated patients with NSCLC and because the percentage of positive expression was often not provided and tested with heterogeneous methods; therefore, we were unable to precisely characterize the interplay between PD-L1 status and HPD. Similarly, tumor mutational burden (TMB) was not available because it was not routinely assessed outside of clinical trials. In our study, only 26 patients (2 classified as having HPD) were treated with immunotherapy in combination with other drugs. Recently, a significant survival benefit for first-line PD-1/PD-L1 inhibitors–chemotherapy (KEYNOTE-021 study,[34] IMpower150 study,[35] KEYNOTE-189 study[3], KEYNOTE 407 study,[36] and IMpower131 study[37]) and PD-1/PD-L1 inhibitor–PD-1/PD-L1 inhibitor (CheckMate 227 study[4]) combinations compared with platinum doublets has been reported. In high-TMB patients with NSCLC, the PFS curves of nivolumab plus ipilimumab and platinum-based chemotherapy treatments cross between 3 and 6 months; in patients receiving platinum-based chemotherapy in combination with pembrolizumab or placebo, an early separation of both PFS and OS curves has been observed.[3,36] These findings suggest a considerable rate of fast progressions or early deaths, potentially due to HPD, in patients treated with double immune checkpoint blockade; in contrast, the addition of chemotherapy to PD-1/PD-L1 inhibitors may hamper PD-1/PD-L1 inhibitor resistance and HPD. Overall, whether HPD is an issue in PD-L1 or TMB selected patients with NSCLC or develops on PD-1/PD-L1 inhibitor combinatorial strategies remains an open question that should be addressed in future studies.

Finally, despite the large sample population herein, it was impossible to define a particular clinical or pathological phenotype for HPD because of the limited number of hyperprogressors. Likewise, the characterization of the molecular basis of HPD remains an unmet need. Some immune checkpoint molecules, such as PD-1 expression[38] and Tim-3 expression,[39] might temper T-regulatory (Treg) cell proliferation and immune suppressive functions, a phenomenon defined as "contra-suppression."[40] Furthermore, a high level of interferon γ (IFN-γ), usually released by PD-1 blockade,[41] may have detrimental effects on immunity as observed in murine mycobacterial infections[42] or in cancer models where increased IFN-γ was associated with activation of tumor immunosuppressive myeloid cells[43] and upregulation of inhibitory metabolites[44] (eg, indoleamine 2,3-dioxygenase) involved in Treg differentiation.[45] Alternatively, PD-1/PD-L1 blockade may upregulate the interleukin 6, interleukin 17, and neutrophil axis, generating a potent aberrant inflammation responsible for immune escape and accelerated growth, as shown in tuberculosis[46] and lung cancer[47] in vivo models. Future studies with prospective assessment of tumor and blood samples from patients with HPD both before treatment and on treatment help clarify the mechanisms behind this phenomenon and its causal relation to treatment.

Conclusions

We identified HPD in 13.8% (56 of 406) of patients with advanced NSCLC treated with PD-1/PD-L1 inhibitors and in only 5.1% (3 of 59) of patients with advanced NSCLC treated with single-agent chemotherapy. In this study, HPD was associated with a high number of metastatic sites at baseline and poor survival (3.4 months), suggesting a detrimental association of immunotherapy in a subset of patients with NSCLC. Additional studies are needed to characterize the molecular basis of HPD.

ARTICLE INFORMATION

Accepted for Publication: May 29, 2018.

Published Online: September 6, 2018. doi:10.1001/jamaoncol.2018.3676

Author Affiliations: Cancer Medicine Department, Gustave Roussy, Villejuif, France (Ferrara, Mezquita, Lahmar, Planchard, Remon, Boucher, Besse); Biostatistics and Epidemiology Department, Gustave Roussy, Villejuif, France (Texier, Koscielny); Pneumology Department, Centre Hospitalier Toulon Sainte-Musse, Toulon, France (Audigier-Valette); Nuclear Medicine Department, Centre Hospitalier Toulon Sainte-Musse, Toulon, France (Tessonnier); Pneumology Department, Centre Hospitalier Universitaire de Toulouse, Université Paul Sabatier, Toulouse, France (Mazieres); Thoracic Oncology Department, Hôpital Bichat-Claude Bernard, Université Paris-Diderot, Paris, France (Zalcman, Brosseau); Medical Oncology Department, Institute Bergonié, Bordeaux, France (Le Moulec, Leroy); Medical Oncology Department, Hôpital Avicenne, Bobigny, France (Duchemann); Service des Maladies Respiratoires, Centre Hospitalier Universitaire de Bordeaux, Bordeaux, France (Lefebvre, Veillon); Pneumology Department, Centre Hospitalier Universitaire de Besançon, Besançon, France (Westeel); Drug Development Department (DITEP), Gustave Roussy, Villejuif, France (Champiat, Ferté, Gazzah, Soria); Pathology Department, Gustave Roussy, Villejuif, France (Adam); Medical Oncology, Fondazione Policlinico Universitario Gemelli IRCCS, Università Cattolica Del Sacro Cuore, Rome, Italy (Bria); Medical Oncology and University of Verona, Verona, Italy (Tortora); Paris-Sud University, Orsay, France (Besse); Radiology Department, Gustave Roussy, Villejuif, France (Caramella).

Author Contributions: Drs Ferrara and Texier had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Besse and Caramella are co-last authors.
Concept and design: Ferrara, Mezquita, Texier, Lahmar, Brosseau, Koscielny, Champiat, Ferté, Bria, Soria, Besse, Caramella.
Acquisition, analysis, or interpretation of data: Ferrara, Mezquita, Texier, Lahmar, Audigier-Valette, Tessonnier, Mazieres, Zalcman, Le Moulec, Leroy, Duchemann, Lefebvre, Veillon, Westeel, Koscielny, Champiat, Ferté, Planchard, Remon, Boucher, Gazzah, Adam, Tortora, Besse, Caramella.
Drafting of the manuscript: Ferrara, Mezquita, Lahmar, Mazieres, Ferté, Remon, Besse.
Critical revision of the manuscript for important intellectual content: Ferrara, Mezquita, Texier, Audigier-Valette, Tessonnier, Mazieres, Zalcman, Brosseau, Le Moulec, Leroy, Duchemann, Lefebvre, Veillon, Westeel, Koscielny, Champiat, Ferté, Planchard, Remon, Boucher, Gazzah, Adam, Bria, Tortora, Soria, Besse, Caramella.
Statistical analysis: Ferrara, Texier, Koscielny, Ferté.
Obtained funding: Mazieres, Le Moulec.
Administrative, technical, or material support: Mezquita, Audigier-Valette, Tessonnier, Veillon, Boucher, Adam.
Supervision: Mezquita, Audigier-Valette, Mazieres, Brosseau, Le Moulec, Champiat, Ferté, Planchard, Remon, Gazzah, Bria, Soria, Besse, Caramella.

Conflict of Interest Disclosures: Dr Audigier-Valette reported serving in a consulting/advisory role for Bristol-Myers Squibb and Merck Sharp & Dohme. Dr Mazieres reported serving in a consulting/advisory role for Bristol-Myers Squibb, Merck Sharp & Dohme, Roche, AstraZeneca, and Eli Lilly and reported receiving research funding from Roche, Bristol-Myers Squibb, and AstraZeneca. Dr Duchemann reported serving in a consulting/advisory role for Roche and Bristol-Myers Squibb and reported receiving travel/accommodation funding from Roche. Dr Westeel reported serving in a consulting/advisory role for Bristol-Myers Squibb, Merck Sharp & Dohme, and AstraZeneca; reported receiving honoraria from Bristol-Myers Squibb and AstraZeneca; reported serving on speakers bureaus for Bristol-Myers Squibb and Merck Sharp & Dohme; and reported receiving travel/accommodation funding from Bristol-Myers Squibb, Roche, and AstraZeneca. Dr Remon reported receiving travel/accommodation funding from Merck Sharp & Dohme and OSE Pharma. Dr Adam reported serving in a consulting/advisory role for Roche, Bristol-Myers Squibb, Merck Sharp & Dohme, and AstraZeneca. Dr Bria reported receiving speakers fees from Bristol-Myers Squibb, Novartis, AstraZeneca, Merck Sharp & Dohme, Celgene, Pfizer, Helsinn, Eli Lilly, and Roche and reported receiving research funding from AstraZeneca, Roche, Open Innovation, Italian Association for Cancer Research (AIRC), and Cariverona Foundation. Dr Soria reported serving in a consulting/advisory role for Roche, AstraZeneca, and Pfizer. Dr Besse reported receiving research funding from GlaxoSmithKline, Roche/Genentech, Clovis Oncology, Pfizer, Boehringer, Eli Lilly, Servier, Onxeo, Bristol-Myers Squibb, Merck Sharp & Dohme, OSE Pharma, Inivata, and AstraZeneca. No other disclosures were reported.

Funding/Support: Dr Ferrara was the recipient of the grant DUERTECC/EURONCO (Diplôme Universitaire Européen de Recherche Translationnelle et Clinique en Cancérologie) for 2016-2017 and of the IASCL (International Association for the Study of Lung Cancer) Young Investigator Award for 2017-2018. Drs Mezquita and Lahmar were recipients of the grant DUERTECC/EURONCO for 2015-2016.

Role of the Funder/Sponsor: DUERTECC/EURONCO provided support for the design and conduct of the study; the collection, analysis, and interpretation of the data; the preparation of the manuscript; and the decision to submit the manuscript for publication.

Meeting Presentations: The preliminary results of this article were presented at the ESMO (European Society for Medical Oncology) 2016 Congress; October 7, 2016; Copenhagen, Denmark, and at the 17th World Conference on Lung Cancer; December 4-7, 2016; Vienna, Austria. The updated results of this article were presented and awarded best poster at the EMSO 2017 Congress; September 10, 2017; Madrid, Spain, and were presented as a mini-oral presentation at the 18th World Conference on Lung Cancer; October 17, 2017; Yokohama, Japan.

Additional Contributions: Sarah MacKenzie, PhD, provided medical writing support funded by Gustave Roussy. No additional compensation was received.

REFERENCES

1. Reck M, Rodríguez-Abreu D, Robinson AG, et al; KEYNOTE-024 Investigators. Pembrolizumab versus chemotherapy for PD-L1–positive non–small-cell lung cancer. *N Engl J Med*. 2016;375(19):1823-1833. doi:10.1056/NEJMoa1606774

2. Brahmer J, Rodríguez-Abreu D, Robinson A, et al. OA 17.06 updated analysis of KEYNOTE-024: pembrolizumab vs platinum-based chemotherapy for advanced NSCLC with PD-L1 TPS ≥50%. *J Thorac Oncol*. 2017;12(11):S1793-S1794. doi:10.1016/j.jtho.2017.09.431

3. Gandhi L, Rodríguez-Abreu D, Gadgeel S, et al; KEYNOTE-189 Investigators. Pembrolizumab plus chemotherapy in metastatic non–small-cell lung cancer. *N Engl J Med*. 2018;378(22):2078-2092. doi:10.1056/NEJMoa1801005

4. Hellmann MD, Ciuleanu T-E, Pluzanski A, et al. Nivolumab plus ipilimumab in lung cancer with a high tumor mutational burden. *N Engl J Med*. 2018;378(22):2093-2104. doi:10.1056/NEJMoa1801946

5. Borghaei H, Paz-Ares L, Horn L, et al. Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. *N Engl J Med*. 2015;373(17):1627-1639. doi:10.1056/NEJMoa1507643

6. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non–small-cell lung cancer. *N Engl J Med*. 2015;373(2):123-135. doi:10.1056/NEJMoa1504627

7. Rittmeyer A, Barlesi F, Waterkamp D, et al; OAK Study Group. Atezolizumab versus docetaxel in patients with previously treated non–small-cell lung cancer (OAK): a phase 3, open-label, multicentre randomised controlled trial. *Lancet*. 2017;389(10066):255-265. doi:10.1016/S0140-6736(16)32517-X

8. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1–positive, advanced non–small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet*. 2016;387(10027):1540-1550. doi:10.1016/S0140-6736(15)01281-7

9. Champiat S, Dercle L, Ammari S, et al. Hyperprogressive disease is a new pattern of progression in cancer patients treated by anti–PD-1/PD-L1. *Clin Cancer Res*. 2017;23(8):1920-1928. doi:10.1158/1078-0432.CCR-16-1741

10. Saâda-Bouzid E, Defaucheux C, Karabajakian A, et al. Hyperprogression during anti-PD-1/PD-L1 therapy in patients with recurrent and/or metastatic head and neck squamous cell carcinoma. *Ann Oncol*. 2017;28(7):1605-1611. doi:10.1093/annonc/mdx178

11. Gomez-Roca C, Koscielny S, Ribrag V, et al. Tumour growth rates and RECIST criteria in early drug development. *Eur J Cancer*. 2011;47(17):2512-2516. doi:10.1016/j.ejca.2011.06.012

12. Seymour L, Bogaerts J, Perrone A, et al; RECIST Working Group. iRECIST: guidelines for response criteria for use in trials testing immunotherapeutics. *Lancet Oncol*. 2017;18(3):e143-e152. doi:10.1016/S1470-2045(17)30074-8

13. Caramella C, Tazdait M, Mezquita L, et al. 1164P: patterns of progression under antiPD1/PDL1 in advanced NSCLC patients allow discriminating pseudo-progression from real progression. *Ann Oncol*. 2017;28(suppl_5). doi:10.1093/annonc/mdx376.029

14. Ferté C, Fernandez M, Hollebecque A, et al. Tumor growth rate is an early indicator of antitumor drug activity in phase I clinical trials. *Clin Cancer Res*. 2014;20(1):246-252. doi:10.1158/1078-0432.CCR-13-2098

15. Giobbie-Hurder A, Gelber RD, Regan MM. Challenges of guarantee-time bias. *J Clin Oncol*. 2013;31(23):2963-2969. doi:10.1200/JCO.2013.49.5283

16. Peters S, Cappuzzo F, Horn L, et al. OA03.05: analysis of early survival in patients with advanced non-squamous NSCLC treated with nivolumab vs docetaxel in CheckMate 057. *J Thorac Oncol*. 2017;12(1):S253. doi:10.1016/j.jtho.2016.11.241

17. Inoue T, Tamiya M, Tamiya A, et al. Analysis of early death in Japanese patients with advanced non-small-cell lung cancer treated with nivolumab. *Clin Lung Cancer*. 2018;19(2):e171-e176. doi:10.1016/j.cllc.2017.09.002

18. Opdivo, INN-nivolumab. http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Product_Information/human/003985/WC500189765.pdf. Accessed September 29, 2017.

19. Mezquita L, Auclin E, Ferrara R, et al. Association of the Lung Immune Prognostic Index with immune checkpoint inhibitor outcomes in patients with advanced non–small cell lung cancer. *JAMA Oncol*. 2018;4(3):351-357. doi:10.1001/jamaoncol.2017.4771

20. Kato S, Goodman A, Walavalkar V, Barkauskas DA, Sharabi A, Kurzrock R. Hyperprogressors after immunotherapy: analysis of genomic alterations associated with accelerated growth rate. *Clin Cancer Res*. 2017;23(15):4242-4250. doi:10.1158/1078-0432.CCR-16-3133

21. Lee CK, Man J, Lord S, et al. Clinical and molecular characteristics associated with survival among patients treated with checkpoint inhibitors for advanced non–small cell lung carcinoma: a systematic review and meta-analysis. *JAMA Oncol*. 2018;4(2):210-216. doi:10.1001/jamaoncol.2017.4427

22. Mellema WW, Burgers SA, Smit EF. Tumor flare after start of RAF inhibition in *KRAS* mutated NSCLC: a case report. *Lung Cancer*. 2015;87(2):201-203. doi:10.1016/j.lungcan.2014.11.014

23. Kuriyama Y, Kim YH, Nagai H, Ozasa H, Sakamori Y, Mishima M. Disease flare after discontinuation of crizotinib in anaplastic lymphoma kinase-positive lung cancer. *Case Rep Oncol*. 2013;6(2):430-433. doi:10.1159/000354756

24. Chaft JE, Oxnard GR, Sima CS, Kris MG, Miller VA, Riely GJ. Disease flare after tyrosine kinase inhibitor discontinuation in patients with *EGFR*-mutant lung cancer and acquired resistance to erlotinib or gefitinib: implications for clinical trial design. *Clin Cancer Res*. 2011;17(19):6298-6303. doi:10.1158/1078-0432.CCR-11-1468

25. Tazdait M, Mezquita L, Lahmar J, et al. Patterns of responses in metastatic NSCLC during PD-1 or PDL-1 inhibitor therapy: comparison of RECIST 1.1, irRECIST and iRECIST criteria. *Eur J Cancer*. 2018;88:38-47. doi:10.1016/j.ejca.2017.10.017

26. Gandara DR, Von Pawel J, Sullivan RN, et al. Impact of atezolizumab (atezo) treatment beyond disease progression (TBP) in advanced NSCLC: results from the randomized phase III OAK study [abstract]. *J Clin Oncol*. 2017;35(15_suppl):9001. doi:10.1200/JCO.2017.35.15_suppl.9001

27. Kazandjian D, Keegan P, Suzman DL, Pazdur R, Blumenthal GM. Characterization of outcomes in patients with metastatic non-small cell lung cancer treated with programmed cell death protein 1 inhibitors past RECIST version 1.1–defined disease progression in clinical trials. *Semin Oncol*. 2017;44(1):3-7. doi:10.1053/j.seminoncol.2017.01.001

28. Long GV, Weber JS, Larkin J, et al. Nivolumab for patients with advanced melanoma treated beyond progression: analysis of 2 phase 3 clinical trials. *JAMA Oncol*. 2017;3(11):1511-1519. doi:10.1001/jamaoncol.2017.1588

29. Hodi FS, Hwu WJ, Kefford R, et al. Evaluation of immune-related response criteria and RECIST v1.1 in patients with advanced melanoma treated with pembrolizumab. *J Clin Oncol*. 2016;34(13):1510-1517. doi:10.1200/JCO.2015.64.0391

30. George S, Motzer RJ, Hammers HJ, et al. Safety and efficacy of nivolumab in patients with metastatic renal cell carcinoma treated beyond progression: a subgroup analysis of a randomized clinical trial. *JAMA Oncol*. 2016;2(9):1179-1186. doi:10.1001/jamaoncol.2016.0775

31. Weinstock C, Maher VE, Zhang L, et al. FDA analysis of treatment beyond disease progression disease (PD) in patients with metastatic renal cell carcinoma (mRCC) treated with nivolumab vs. everolimus [abstract]. *J Clin Oncol*. 2016;34(15_suppl):4508. doi:10.1200/JCO.2016.34.15_suppl.4508

32. Escudier B, Motzer RJ, Sharma P, et al. Treatment beyond progression in patients with advanced renal cell carcinoma treated with nivolumab in CheckMate 025. *Eur Urol*. 2017;72(3):368-376. doi:10.1016/j.eururo.2017.03.037

33. Blumenthal GM, Theoret MR, Pazdur R. Treatment beyond progression with immune checkpoint inhibitors: known unknowns. *JAMA Oncol*. 2017;3(11):1473-1474. doi:10.1001/jamaoncol.2017.1819

34. Langer CJ, Gadgeel SM, Borghaei H, et al; KEYNOTE-021 Investigators. Carboplatin and pemetrexed with or without pembrolizumab for advanced, non-squamous non-small-cell lung cancer: a randomised, phase 2 cohort of the open-label KEYNOTE-021 study. *Lancet Oncol*. 2016;17(11):1497-1508. doi:10.1016/S1470-2045(16)30498-3

35. Socinski MA, Jotte RM, Cappuzzo F, et al. Atezolizumab for first-line treatment of metastatic nonsquamous NSCLC. *N Engl J Med*. 2018;378(24):2288-2301. doi:10.1056/NEJMoa1716948

36. Paz-Ares L, Luft A, Tafreshi A, et al. Phase 3 study of carboplatin-paclitaxel/nab-paclitaxel (Chemo) with or without pembrolizumab (Pembro) for patients (Pts) with metastatic squamous (Sq) non-small cell lung cancer (NSCLC). *J Clin Oncol*. 2018;36(15_suppl):105. doi:10.1200/JCO.2018.36.15

37. Jotte R, Cappuzzo F, Vynnychenko I, et al. IMpower131: Primary PFS and safety analysis of a randomized phase III study of atezolizumab + carboplatin + paclitaxel or nab-paclitaxel vs carboplatin + nab-paclitaxel as 1L therapy in advanced squamous NSCLC. *J Clin Oncol*. 2018;36(18_suppl):LBA9000. doi:10.1200/JCO.2018.36.18_suppl.LBA9000

38. Franceschini D, Paroli M, Francavilla V, et al. PD-L1 negatively regulates $CD4^+CD25^+Foxp3^+$ Tregs by limiting STAT-5 phosphorylation in patients chronically infected with HCV. *J Clin Invest*. 2009;119(3):551-564. doi:10.1172/JCI36604

39. Moorman JP, Wang JM, Zhang Y, et al. Tim-3 pathway controls regulatory and effector T cell balance during hepatitis C virus infection. *J Immunol*. 2012;189(2):755-766. doi:10.4049/jimmunol.1200162

40. Barnaba V, Schinzari V. Induction, control, and plasticity of Treg cells: the immune regulatory network revised? *Eur J Immunol*. 2013;43(2):318-322. doi:10.1002/eji.201243265

41. Peng W, Liu C, Xu C, et al. PD-1 blockade enhances T-cell migration to tumors by elevating IFN-γ inducible chemokines. *Cancer Res*. 2012;72(20):5209-5218. doi:10.1158/0008-5472.CAN-12-1187

42. Sakai S, Kauffman KD, Sallin MA, et al. CD4 T cell–derived IFN-γ plays a minimal role in control of pulmonary *Mycobacterium tuberculosis* infection and must be actively repressed by PD-1 to prevent lethal disease. *PLoS Pathog*. 2016;12(5):e1005667. doi:10.1371/journal.ppat.1005667

43. Huang B, Pan PY, Li Q, et al. Gr-1^+CD115^+ immature myeloid suppressor cells mediate the development of tumor-induced T regulatory cells and T-cell anergy in tumor-bearing host. *Cancer Res*. 2006;66(2):1123-1131. doi:10.1158/0008-5472.CAN-05-1299

44. Spranger S, Spaapen RM, Zha Y, et al. Up-regulation of PD-L1, IDO, and T_{regs} in the melanoma tumor microenvironment is driven by $CD8^+$ T cells. *Sci Transl Med*. 2013;5(200):200ra116. doi:10.1126/scitranslmed.3006504

45. Baban B, Chandler PR, Sharma MD, et al. IDO activates regulatory T cells and blocks their conversion into T_H17-like T cells. *J Immunol*. 2009;183(4):2475-2483. doi:10.4049/jimmunol.0900986

46. Lázár-Molnár E, Chen B, Sweeney KA, et al. Programmed death-1 (PD-1)–deficient mice are extraordinarily sensitive to tuberculosis. *Proc Natl Acad Sci U S A*. 2010;107(30):13402-13407. doi:10.1073/pnas.1007394107

47. Akbay EA, Koyama S, Liu Y, et al. Interleukin-17A promotes lung tumor progression through neutrophil attraction to tumor sites and mediating resistance to PD-1 blockade. *J Thorac Oncol*. 2017;12(8):1268-1279. doi:10.1016/j.jtho.2017.04.017

JAMA Oncology | Original Investigation

Clarification of Definitions of Hyperprogressive Disease During Immunotherapy for Non–Small Cell Lung Cancer

Baptiste Kas, BSc; Hugues Talbot, PhD; Roberto Ferrara, MD, PhD; Colombe Richard; Jean-Philippe Lamarque, MSc; Stéphanie Pitre-Champagnat, PhD; David Planchard, MD, PhD; Corinne Balleyguier, MD, PhD; Benjamin Besse, MD, PhD; Laura Mezquita, MD, PhD; Nathalie Lassau, MD, PhD; Caroline Caramella, MD

IMPORTANCE Hyperprogressive disease (HPD) is an aggressive pattern of progression reported for patients treated with programmed cell death 1 (PD-1)/programmed cell death 1 ligand (PD-L1) inhibitors as a single agent in several studies. However, the use of different definitions to assess this phenomenon induces the risk of describing different tumoral behaviors.

OBJECTIVE To assess the accuracy of each HPD definition to identify the frequency of HPD and the association with poorer outcomes of immune-checkpoint inhibitor (ICI) treatment in patients with advanced non–small cell lung cancer (NSCLC) and to provide an optimized and homogenized definition based on all previous criteria for identifying HPD.

DESIGN, SETTING, AND PARTICIPANTS This retrospective cohort study included 406 patients with advanced NSCLC treated with PD-1/PD-L1 inhibitors from November 1, 2012, to April 5, 2017, in 8 French institutions. Measurable lesions were defined using the Response Evaluation Criteria in Solid Tumours (RECIST) 1.1 criteria on at least 2 computed tomographic scans before the initiation of ICI therapy and 1 computed tomographic scan during treatment. Data were analyzed from November 1, 2012, to August 1, 2019.

INTERVENTIONS All dynamic indexes used in the previous proposed definitions, such as the tumor growth rate (TGR) or tumor growth kinetics (TGK), were calculated before and during treatment.

RESULTS Among the 406 patients with NSCLC included in the analysis (259 male [63.8%]; median age at start of ICI treatment, 64 [range, 30-91] years), the different definitions resulted in incidences of the HPD phenomenon varying from 5.4% (n = 22) to 18.5% (n = 75). The concordance between these different definitions (using the Jaccard similarity index) varied from 33.3% to 69.3%. For every definition, HPD was associated with poorer survival (range of median overall survival, 3.4 [95% CI, 1.9-8.4] to 6.0 [95% CI, 3.7-9.4] months). The difference between TGR before and during therapy (ΔTGR) was the most correlated with poor overall survival with an initial plateau for a larger number of patients and a slower increase, and it had the highest ability to distinguish patients with HPD from those with PD not classified as HPD. In addition, an optimal threshold of ΔTGR of greater than 100 was identified for this distinction.

CONCLUSIONS AND RELEVANCE These findings suggest that the previous 5 definitions of HPD are not representative of the same tumor behavior. A new definition, based on ΔTGR of greater than 100, appears to be more relevant to describe the characteristics expected with HPD (increase of the tumor kinetics and poor survival) in this cohort of patients with NSCLC. Additional studies on larger groups of patients are necessary to confirm the accuracy and validate our proposed definition.

Invited Commentary

Supplemental content

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Caroline Caramella, MD, Radiology Department, Gustave Roussy, Université Paris-Saclay, Villejuif, France, 114 Rue Édouard-Vaillant, 94805 Villejuif Cedex, France (caroline.caramella@gustaveroussy.fr).

JAMA Oncol. doi:10.1001/jamaoncol.2020.1634
Published online June 11, 2020.

Immune-checkpoint inhibitors (ICIs) have been one of the major developments in cancer therapy in the past decade and are approved for various tumor types, such as melanoma, non-small cell lung cancer (NSCLC), renal cell carcinoma, or head and neck squamous cell carcinoma.[1-4] Their approval has caused a revolution in the therapeutic approach, because they differ from conventional cytotoxic treatments and molecularly targeted agents in their mechanism and in the response patterns with which they are associated.

Immune-checkpoint inhibitors have indeed demonstrated survival benefits, including complete and partial responses as well as long-term remission, and have been associated with novel patterns of responses such as pseudoprogression, defined as an initial increase in the tumor burden followed by a later or a dissociated response.[5] Of greater concern, several studies have reported a possible deleterious effect of ICI in a subpopulation of patients, with dramatic tumor growth following the initiation of the therapy, termed *hyperprogressive disease* (HPD).[6] This phenomenon is clinically defined as an unexpected acceleration of the tumor kinetics that can be measured on imaging with dynamic parameters.

To define and quantify the incidence of this phenomenon, parameters have been used that include tumor growth rate (TGR),[7] tumor growth kinetics (TGK),[8] and time to treatment failure.[9] However, to date, no consensual HPD definition has been proposed, and the risk of describing different tumoral behaviors exists. The objectives of the present study are to achieve an advanced comprehensive comparison of the different definitions of HPD in a cohort of patients with NSCLC to evaluate the association of the definition with the related incidence and the HPD subset constitution and the association between each HPD definition and overall survival.

Key Points

Question Are the different definitions used to assess hyperprogressive disease representative of the same tumoral behavior?

Findings For this multicenter cohort study of 406 patients with advanced non-small cell lung cancer treated with programmed cell death 1/programmed cell death 1 ligand inhibitors, the 5 disease definitions assessed resulted in diverse incidences, different patient characteristics, and different associations with survival outcomes. A new criterion of difference in tumor growth rate of greater than 100 showed more accuracy in assessing hyperprogressive disease.

Meaning These findings suggest that the 5 definitions assessed are not representative of the same tumoral behavior.

Methods

Patients

This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline. This study was approved by the institutional review board of Gustave Roussy, which did not require informed consent from participants because of the retrospective nature. Data from patients with NSCLC who were treated with programmed cell death 1 (PD-1)/programmed cell death 1 ligand (PD-L1) inhibitors from November 1, 2012, to April 5, 2017, in 8 French institutions were retrospectively collected and analyzed. Patients with confirmed stage III or IV NSCLC and for whom computed tomographic (CT) scans were available for radiological evaluation were included. Most patients received prior treatments (eTable in the Supplement).

At least 2 CT scans before the beginning of the ICI therapy and 1 CT scan during the treatment were required for assessing the dynamic indexes. The baseline CT scan was performed in the 6 weeks preceding the initiation of ICI treatment, and a minimum of 2 weeks between different CT evaluations was expected, resulting in the inclusion of 406 patients in the final cohort (eFigure 1 in the Supplement). The median interval from prebaseline to baseline CT scans was 60 days (range, 14-120 days [2 weeks to 4 months]); from baseline to follow-up CT scans, 62 days (range, 14-120 days [2 weeks to 3 months]).

Definitions of Tumor Dynamics

In the recent literature, HPD has been defined in different studies using 5 different criteria as already emphasized in the study by Fuentes-Antrás et al.[10] A summary of existing definitions is available in **Table 1**. For a better understanding of our analysis, a distinction between the terms *definition* and *index* has

Table 1. Main Different Definitions of HPD According to Previous Studies

Component	Source				
	Champiat et al,[6] November 2016	Kato et al,[9] March 2017	Saâda-Bouzid et al,[11] April 2017	Singavi et al,[12] September 2017	Ferrara et al,[13] November 2018
Letter	A	B	C	D	E
Definition[a]	RECIST progression and TGR-exp/TGR-ref >2	TTF <2 mo, RECIST>50%, and progression pace >2-fold	TGK-exp/TGK-ref >2	RECIST progression, RECIST >50%, and TGR-exp/TGR-ref >2	RECIST progression and TGR-exp − TGR-ref >50
Reported HPD incidence, No. (%)	12/131 (9.2)	6/155 (3.9)	10/34 (29.4)	5 patients	56/406 (13.8)
Histologic types	Various: melanoma, 34%; lung, 10%	Various: melanoma, 33%; NSCLC, 25%	Head and neck squamous cell carcinoma	Various	NSCLC

Abbreviations: HPD, hyperprogressive disease; NSCLC, non-small cell lung carcinoma; RECIST, Response Evaluation Criteria in Solid Tumours; TGK, tumor growth kinetics; TGR, tumor growth rate; TTF, time to treatment failure.

[a] Indexes are calculated using TGR and TGK values before (ref) and during (exp) therapy.

been established, with *index* referring to the mathematical parameters, such as TGR or TGK, combined and used with thresholds to define HPD.

Tumor growth rate is defined by Gomez-Roca et al[7] as the percentage increase in tumor volume per month, following TGR = 100 [exp(TG) - 1], where TG is 3-log (S_t/S_O) and where S_t and S_O are the tumor sizes at times t and 0, defined as the sum of the longest diameters of the target lesions as per the Response Evaluation Criteria in Solid Tumours (RECIST) 1.1 criteria.[14] Tumor growth kinetics are defined as the change in the tumor size per unit of time (in mm/d) as reported by Le Tourneau et al[8]: TGK = $(S_t - S_O)/(t - t_O)$. Both indexes were calculated before and during treatment to evaluate any change in the tumor kinetics. For both indexes and for every patient, the RECIST sum was computed with the target lesions defined at baseline of ICI. For an easier reading, each of these definitions will be named using letters from A to E as in Table 1. Of these 5 definitions, 3 (A, D, and E) rest on the hypothesis of a natural exponential growth of the tumor volume with time. Assuming that at time t, the volume V_t can be expressed as $V_t = V_O \exp(TG \times t)$, this hypothesis leads directly to the use of TGR as the HPD index. Nevertheless, these 3 definitions are not strictly equivalent.

Champiat et al[6] (definition A) defined HPD as at least a 2-fold increase of the TGR from before (ref) to during (exp) therapy (TGR-exp:TGR-ref ≥2). In other words, patients with HPD are characterized by a twice higher percentage increase in volume per month during immunotherapy than before. A later study by Singavi et al[12] (definition D) takes the same definition as A and adds a condition on the RECIST percentage increase during ICI treatment of more than 50%. The study by Ferrara et al[13] (definition E) assumes that HPD is characterized by a difference (and not a ratio) superior to 50% between TGR-exp and TGR-ref (ΔTGR), suggesting that the increase in volume per month during ICI therapy must be 50% higher than that expected with the increase before treatment.

Saâda-Bouzid et al[11] (definition C) relied on the use of TGK, which does not account for the hypothesis of the natural exponential growth of the tumor and uses the diameters rather than the volume. Finally, Kato et al[9] (definition B) defined HPD using 3 conditions: tumor kinetics (progression pace >2-fold), the RECIST percentage (increase in the tumor burden during ICI >50%), and ongoing treatment (time to treatment failure, ≤2 months). In the analysis, tumor burden at prebaseline, baseline, and postbaseline evaluations will be termed *S-prebaseline*, *S-baseline*, and *S-postbaseline*, respectively.

Statistical Analysis

Data were analyzed from November 1, 2012, to August 1, 2019. Concordance between the constitutions of the HPD groups for the different definitions was evaluated using the similarity Jaccard index. The influence of each definition on the designation of HPD was then theoretically analyzed as a function of the RECIST percentage before immunotherapy and the RECIST percentage during immunotherapy. To do so, we represented the mathematical criteria (TGK-exp:TGK-ref >2, TGR-exp:TGR-ref >2, and ΔTGR>50) under the form of isolines and compared the respective positions of the curves. For an identical RECIST percentage before immunotherapy, the curve corresponds to the definition that requires a higher RECIST percentage during therapy to assess HPD and that is therefore more restrictive.

To investigate the association between HPD status and survival outcomes, patients with initial PD as defined per RECIST 1.1 were divided into those with HPD and those with PD without HPD. Overall survival was defined as the time from the initiation of the ICI therapy to the death of the patient due to any cause; time to treatment failure, as the duration from the beginning to the discontinuation of the treatment for any reason, including toxicity, progression, patient's choice, or death. Landmark survival analyses (eFigure 2 and 3 in the Supplement) were performed using the Kaplan-Meier method,[15] and the log-rank test was used for comparison. Two-sided $P < .05$ was considered statistically significant.

To evaluate the prognostic value of each index, we divided the patients with PD into 2 groups: those with the highest values of the index (N) and those with other PD. We then computed the median overall survival as a function of the number N. Studying the influence of N on the landmark analysis, our objective was to determine which indexes showed a clear ability to distinguish patients with both acceleration of the tumor growth and poor survival and thereby determine which index and threshold led to the most significant distinction between the groups. All the statistical and mathematical analyses were performed using Python software, version 3.0 (Python Software Foundation, Python Language Reference).

Results

Of 406 patients included in the analyses, 259 (63.8%) were male and 147 (36.2%) were female. The median age at the beginning of ICI was 64 (range, 30-91) years. Of these, 207 were assessed as having PD while receiving ICI therapy at first evaluation, owing to an increase in target lesions of greater than 20% in 143 and an increase in nontarget lesions or appearance in 64. Nineteen patients with PD were retrospectively assessed as having pseudoprogression and were excluded from the analysis. The complete characteristics of the patients have previously been described by Ferrara et al.[13]

Influence of the Definition on Incidence

When we applied the different definitions of HPD to the 406 patients in the cohort analysis, HPD incidence varied from 22 (5.4%) to 75 (18.5%) patients. Definition B of HPD appeared to be the most restrictive definition, with the smallest incidence (5.4%) (**Table 2**).

HPD Subsets With the Different HPD Definitions

Nineteen patients underwent assessment of HPD by all 5 definitions (A, B, C, D, and E). The maximum value of the similarity index was 69.3%, reached for definitions A and C; and the minimum value was 33.3%, reached for the definitions of C and D (**Table 3**), demonstrating different HPD subpopulations with each definition. To understand which patients are classified as having HPD by each definition, the characteristics of pa-

Table 2. Incidence and Overall Survival of Patients With HPD According to the Different Definitions

Variable	Definition A	Definition B	Definition C	Definition D	Definition E
Incidence, No./total No. (%)	52/406 (12.8)	22/406 (5.4)	75/406 (18.5)	25/406 (6.2)	56/406 (13.8)
Median OS (95% CI), mo					
HPD group	5.1 (3.3-9.4)	3.4 (1.9-8.4)	6.0 (3.7-9.4)	5.1 (3.2-10.3)	4.0 (2.9-7.5)
PD non-HPD group	6.3 (5.0-7.5)	6.4 (5.6-8.4)	6.2 (5.0-7.5)	6.2 (5.3-7.5)	6.4 (5.5-8.4)
P value[a]	.45	<.001	.62	.59	.14

Abbreviations: HPD, hyperprogressive disease (PD); OS, overall survival.

[a] Log-rank tests were computed to test the significance of this distinction.

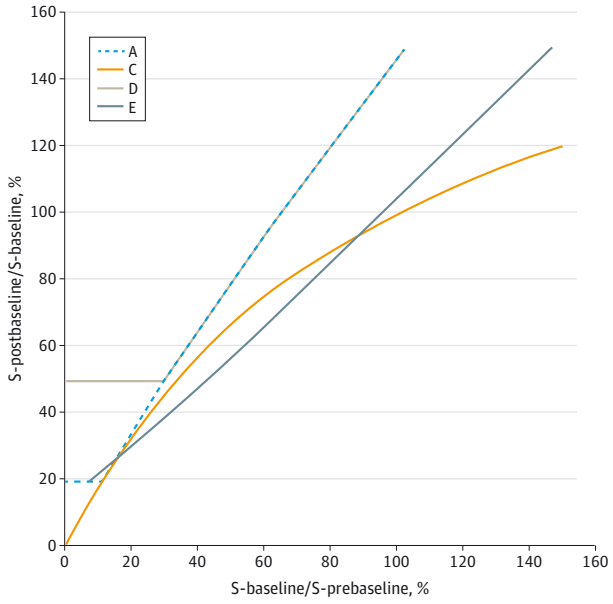
Table 3. Influence of the Definition on the HPD Subset Constitution

	Jaccard index value, No. (%)[a]				
Definition	Definition A (n = 52)	Definition B (n = 22)	Definition C (n = 75)	Definition D (n = 25)	Definition E (n = 56)
A	NA	NA	NA	NA	NA
B	19 (34.5)	NA	NA	NA	NA
C	52 (69.3)	21 (27.6)	NA	NA	NA
D	25 (48.1)	22 (67.9)	25 (33.3)	NA	NA
E	44 (68.8)	23 (34.8)	49 (59.8)	24 (47.4)	NA

Abbreviations: HPD, hyperprogressive disease; NA, not applicable.

[a] The Jaccard index value corresponds to the proportion of patients with HPD in common among all patients assessed as having HPD by one definition or the other. For instance, definitions A and E assessed 44 patients with HPD in common but another 20 patients were assessed as having HPD by only 1 of the 2 definitions (52 − 44 = 8 in A and 56 − 44 = 12 in E), leading to a similarity index of 44 of 64 = 68.8%.

Figure 1. Areas of Hyperprogressive Disease (HPD) Incidence According to Response Evaluation Criteria in Solid Tumours (RECIST)



Data are shown before (x-axis) and during (y-axis) programmed cell death 1/programmed cell death 1 ligand therapy for S-baseline/S-prebaseline of greater than 1, where S indicates tumor size. Isolines for the definitions mark the frontier between patients with progressive disease and those with HPD. For each definition, patients in the area above are assessed as having HPD, whereas patients in the area below are not. These isolines were drawn in the ideal case of a period of 1 month between 2 computed tomographic scans. Definition B does not only rely on considerations of the tumor size and kinetics and therefore could not be included in this comparison.

tients for definitions A, C, D, and E as regarding the RECIST percentages before and during ICI therapy have been modeled in **Figure 1.**

Two cases were distinguished. In S-baseline/S-prebaseline of greater than 1, for a pretherapy increase of the target lesions' size, definitions based on TGR or TGK tended to associate HPD with high values of the RECIST percentage during ICI therapy (above the line in Figure 1). However, the corresponding curves did not overlap, and different situations must be distinguished. Definition C is mathematically more likely to diagnose HPD among patients with a pretreatment increase from 1% to 15% (stable disease according to RECIST criteria) and among patients with a high pretreatment progression of greater than 90%.

Definition A requires the highest RECIST percentage during therapy to define patients with HPD. Definition D adds to definition A the condition RECIST percentage of greater than 50% and is even more restrictive until a pretreatment increase of 40%, when both curves overlap. Finally, definition E tends to diagnose more HPD among patients with a pretreatment progression with a RECIST percentage from 15% to 90% than with other definitions.

For a pretherapy decrease of the RECIST sum, the difference between the definitions is even more substantial. Using mathematical ratios, such as definitions A, B, and D, no patient with S-baseline/S-prebaseline of less than 1 can ever be considered to have HPD; the 3 conditions S-baseline/S-prebaseline of less than 1, S-postbaseline/S-baseline of greater than 1.2, and TGR ratio of greater than 2 (or TGK ratio >2) cannot be satisfied simultaneously. However, using a subtraction, such as definitions C and E, allows patients with a small S-baseline/S-prebaseline and a high S-postbaseline/S-baseline to be assessed as having HPD. These patients were nonetheless declared to have PD during pretreatment because of the appearance of new lesions only, which are accounted for in none of the definitions.

Association Between HPD Definitions and OS

For each of the 5 definitions, the landmark survival analysis emphasized a worse outcome for the patients with HPD compared with the patients with PD. The median overall survival for patients with HPD, which ranged from 3.4 (95% CI, 1.9-8.4) to 6.0 (95% CI, 3.7-9.4) months, was smaller than that for patients without HPD, which ranged from 6.2 (95% CI, 5.0-7.5) to 6.4 (95% CI, 5.6-8.4) months. The gap between the median overall survival of the 2 groups varied from 0.2 months (definition C) to 3.0 months (definition B), thus highlighting a disparity in the correlation of the different HPD definitions with outcomes (Table 2).

Log-rank tests were computed to test the statistical significance of the differentiation between PD with and without HPD. Only definition B demonstrated a significant distinction (3.0 months; $P < .001$) (Table 2).

Prognostic Value of the Different Indexes

To study the prognostic values of the different indexes (TGK-exp:TGK-ref, TGR-exp:TGR-ref, and ΔTGR) on the 2 groups gathering the highest values on the index (N) and the 207 - N other values on the index (207 patients had PD at first evaluation), we obtained the curves shown in **Figure 2**. For all 3 indexes, the highest values (ie, the highest increases of TGK during therapy) tended to be associated with the poorest survival outcomes. However, the curves of Figure 2 showed differences in the size of the population associated with these characteristics and in the amplitude and the stability of the gap between the median overall survival of both groups.

Figure 2B-D describes the testing of each specific index. Each time, the objective was to determine how efficiently each index discriminated between patients with HPD and those with PD without HPD. In other words, how different are the overall survival curves for both groups? Each time, N = 1 is equivalent to the patient with the most HPD according to the chosen index and N = 2, the 2 patients with the most HPD, etc. When the number N is small, the few patients with the most HPD are assessed as having HPD, and the median overall survival of the HPD group should be much smaller than that of the non-HPD group. When the number N is larger, many patients are assessed as having HPD, which means that the overall survival of the HPD group should be closer to that of the non-HPD group. Therefore, a good index is associated with a large gap between the curves for small N values. The wider and the longer the gap between curves for both groups, the better the index is able to discriminate between patients with HPD and those with PD without HPD. The curves of TGR-exp:TGR-ref and TGK-exp:TGK-ref (Figure 2B and D) appeared to be similar, highlighting a similar distribution of the values among patients (ie, ranking patients according to their values of TGR-exp:TGR-ref or TGK-exp:TGK-ref leads to a close result). Both curves showed an initial plateau until 20 patients but emphasized an important instability followed by a sharp increase of the median overall survival that even overtakes that of patients with PD for a larger number N.

The curve of ΔTGR (Figure 2C) also revealed an initial plateau for a larger number of 40 to 50 patients and a slower increase, demonstrating a greater correlation with overall survival. To confirm the relevance of these distinctions between the 2 groups, we further investigated whether a log-rank test was significant for the different indexes and the different thresholds N. All 3 indexes locally reached a significant $P < .05$ for a small N. However, only ΔTGR remained at $P < .05$ for a range of N, with P values for both other indexes oscillating between significant and nonsignificant values while N increased. The P value for ΔTGR remained significant until a maximum N of 34 patients in the first group corresponding to a threshold ΔTGR of greater than 102.

Discussion

Previous studies[6,9-13] on HPD during immunotherapy reported different incidences of the phenomenon, varying from 4% to 29%. The causes for such a disparity might include the diversity in cancer histology as well as the size and source of the study cohort constitution (Table 1). However, as already emphasized by Kim et al,[16] the metrics used for HPD assessment could also be a major explanation for this inconsistency. To our knowledge, our study is the first one to offer a detailed analytical comparison of all the definitions that have been used so far to assess this phenomenon.

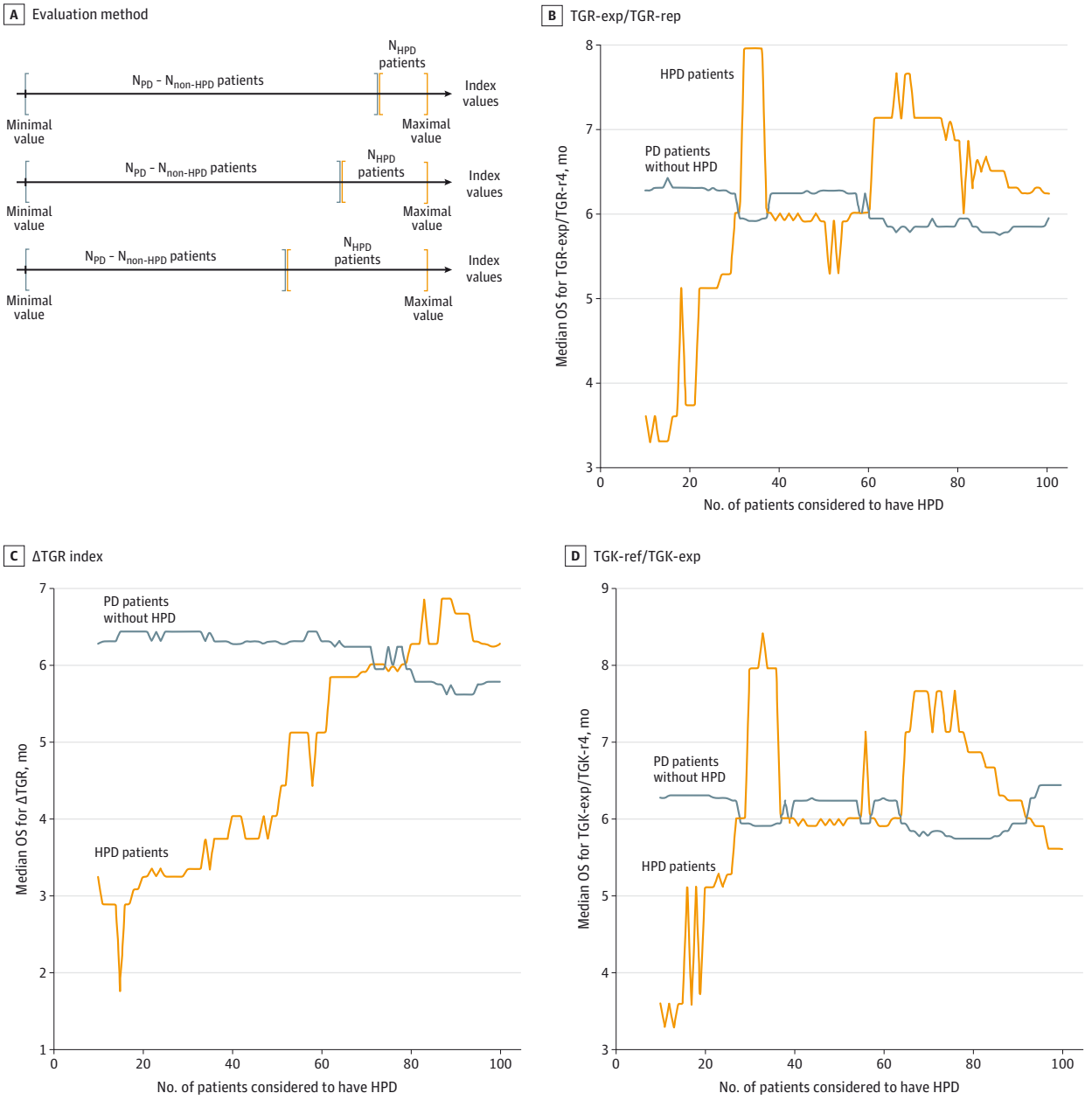
In this study, the rates of HPD with the different definitions applied to the same NSCLC cohort appeared to be concordant with the previously reported studies, with the exception of definition C from Saâda-Bouzid[11] that showed a smaller incidence of patients with HPD in our cohort (18.5% vs 29%). No reasoned explanation can be given for such a gap at this stage, but the effect of the patients' characteristics and the histologic findings (NSCLC vs previously irradiated squamous cell carcinoma of the head and neck) on such a result should be further analyzed.

The results of the present study first and foremost point out the high disparity in HPD incidences due to the definitions themselves, with a number of patients with HPD that can vary from 1 to 4 in the same cohort (22 patients for definition B compared with 75 patients for definition C). The choice of the definition seems therefore to be a major reason for the diversity observed among previous studies. Beyond the question of incidences, these results also highlight the fact that the groups of patients with HPD appear to differ from one definition to the other, with only 19 patients being common to all definitions. More precisely, the similarity measures show that the so-called HPD for different definitions is not representative of the same tumoral behavior.

Consequently, the definitions do not correlate in the same way with overall survival. Most of them proved no ability to establish a clear distinction between the overall survival of patients with HPD compared with PD without HPD. Indeed, only 2 definitions appeared to be (statistically) significantly correlated with a worse overall survival for patients with HPD. This result should be moderated by the fact that the small overall survival is itself a criterion taken into account in both these definitions.

Trying to extract thresholds to align the patients with the worst progression and the worst survival prognosis, we showed

Figure 2. Evaluation of the Prognostic Value of Each Index



A, Evaluation method of the prognostic value of each index. For each index, patients were categorized into 2 groups according to their index value: those with the highest value (N) and those with progressive disease (N_{PD}) minus N patients, to study the difference between the N with worst progression and the N_{PD} – N other patients. We first set N = 10 and then increased its value until it reached half the number of patients with PD. This determines for which N the difference becomes negligible, which is the limit between patients with HPD and PD. A comparative landmark analysis was performed on these 2 groups for each value of N. B, Median overall survival (OS) curves for patients with HPD compared with PD without HPD according to the threshold ratio of tumor growth rate before (ref) and during (exp) therapy (TGR-exp:TGR-ref index) as a function of the number N of patients considered to have HPD. C, Median OS curves for HPD compared with PD without HPD according to the threshold TGR-exp – TGR-ref (ΔTGR) index as a function of the number N of patients considered to have HPD. D, Median OS curves for HPD compared with PD without HPD according to the threshold tumor growth kinetics before (ref) and during (exp) therapy (TGK-exp:TGK-ref) index as a function of the number N of patients considered to have HPD.

that only the index ΔTGR appeared to be likely to distinguish subsets of patients with the characteristics expected with HPD status: acceleration of the tumor growth combined with a poor overall survival. For this index, the significance of the distinction between patients with HPD and those with PD without HPD in terms of median overall survival was reached for 34 patients (8.4%) less than in the previous studies using ΔTGR and corresponding to an approximate threshold ΔTGR of greater than 100. Therefore, to be in accordance with the concept of HPD that assumes both a high increase of the tumor kinetics

and a poor survival outcome, the following definition based on our cohort could appear more relevant: RECIST percentage during therapy of greater than 20% (S-postbaseline/S-baseline >1.2) and ΔTGR of greater than 100.

Limitations

Some limitations to our study should be noted. First, whereas HPD behavior was initially evaluated in a mixed oncologic population[6,9] or in patients with head and neck squamous cell carcinoma,[11] our model only includes patients with NSCLC. Studies with larger groups of patients with different characteristics would be necessary to confirm the accuracy of our definition of HPD that was empirically determined.

Second, the characterization of HPD remains difficult on a routine basis, because pretreatment imaging is required. Moreover, all definitions of HPD are based on the measurement of target lesions following RECIST 1.1 criteria, thus not accounting for the unequivocal progression of nontarget lesions or the appearance of a new lesion. This bias, as well as the artificial exclusion of patients who died before having the requisite posttreatment imaging, making it impossible to associate the death with HPD, may lead to an underestimation of the phenomenon.

Patients with 2-month intervals between CT scans were selected when possible. A more precise timeframe would be ideal for improved precision of HPD assessment, and a larger prospective database would be necessary in further studies.

Conclusions

In this study, we observed the existence of an increase of kinetics of some patients with NSCLC receiving ICI, which we define as HPD. We demonstrated highly variable HPD rates, based on the 5 previous definitions of HPD, and thus suggest an optimized definition: greater than 20% RECIST and ΔTGR of greater than 100, correlated with poor IO outcomes. A biological explanation and surrogate are urgently needed to identify as soon as possible patients with HPD receiving ICI.

ARTICLE INFORMATION

Accepted for Publication: March 30, 2020.

Published Online: June 11, 2020. doi:10.1001/jamaoncol.2020.1634

Author Affiliations: UMR (Unité Mixte de Recherche) 1281, Université Paris-Saclay, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Commissariat à l'énergie Atomique et Aux Énergies Alternatives, Laboratoire d'Imagerie Biomédicale Multimodale Paris-Saclay, Villejuif, France (Kas, Richard, Lamarque, Pitre-Champagnat, Balleyguier, Lassau, Caramella); Center for Visual Computing, CentraleSupelec, Inria, Université Paris-Saclay, Gif-sur-Yvette, France (Talbot); Department of Medical Oncology, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico Istituto Nazionale dei Tumori, Milan, Italy (Ferrara); Cancer Medicine Department, Gustave Roussy, Université Paris-Saclay, Villejuif, France (Planchard, Besse, Mezquita); Radiology Department, Gustave Roussy, Université Paris-Saclay, Villejuif, France (Balleyguier, Lassau, Caramella); Medical Oncology Department, Hospital Clínic, Barcelona, Spain (Mezquita); Translational Genomics and Targeted Therapeutics in Solid Tumours, August Pi i Sunyer Biomedical Research Institute, Barcelona, Spain (Mezquita).

Author Contributions: Dr Caramella had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Concept and design: Kas, Ferrara, Pitre-Champagnat, Balleyguier, Lassau, Caramella.
Acquisition, analysis, or interpretation of data: Kas, Talbot, Richard, Lamarque, Planchard, Besse, Mezquita, Caramella.
Drafting of the manuscript: Kas, Talbot, Ferrara, Richard, Mezquita, Caramella.
Critical revision of the manuscript for important intellectual content: Kas, Talbot, Ferrara, Lamarque, Planchard, Pitre-Champagnat, Balleyguier, Besse, Mezquita, Lassau, Caramella.
Statistical analysis: Kas, Talbot, Richard, Lassau, Caramella.
Obtained funding: Lassau.
Administrative, technical, or material support: Richard, Lamarque, Mezquita, Lassau, Caramella.
Supervision: Kas, Talbot, Ferrara, Planchard, Pitre-Champagnat, Balleyguier, Besse, Lassau, Caramella.

Conflict of Interest Disclosures: Dr Talbot reported performing consulting, advisory role, or lectures for GE Healthcare, Heartflow, Inc, and Sanofi SA and receiving honoraria from Qualia Systems and WeDiagnostiX. Dr Ferrara reported receiving personal fees Merck Sharp & Dohme. Dr Planchard reported performing consulting, advisory role, or lectures for AstraZeneca plc, Bristol-Myers Squibb, Boehringer Ingelheim, Celgene Corporation, Daiichi Sankyo Company, Limited, Eli Lilly and Company, Merck & Co, Novartis International AG, Pfizer, Inc, prIME Oncology, PeerCME, and Roche Diagnostics; honoraria from AstraZeneca plc, Bristol-Myers Squibb, Boehringer Ingelheim, Celgene Corporation, Eli Lilly and Company, Merck & Co, Novartis International AG, Pfizer, Inc, prIME Oncology, PeerCME, and Roche Diagnostics; clinical trials research for AstraZeneca plc, Bristol-Myers Squibb, Boehringer Ingelheim, Eli Lilly and Company, Merck & Co, Novartis International AG, Pfizer, Inc, Roche Diagnositcs, MedImmune, LLC, Sanofi-Aventis, Taiho Pharmaceutical Co, Ltd, NovoCure Limited, and Daiichi Sankyo Company, Limited; and travel, accommodations, and/or expenses from AstraZeneca plc, Roche Diagnositcs, Novartis International AG, prIME Oncology, and Pfizer, Inc. Dr Besse reported performing sponsored research at Gustave Roussy Cancer Center, AbbVie, Inc, Amgen, Inc, AstraZeneca plc, Biogen, Inc, Blueprint Medicines Corporation, Bristol-Myers Squibb, Celgene Corporatoin, Eli Lilly and Company, GlaxoSmithKline plc, Ignyta, Inc, Ipsen, Merck KGaA, Merck Sharp & Dohme, Nektar Therapeutics, Onxeo, Pfizer, Inc, PharmaMar, Sanofi SA, Spectrum Pharmaceuticals, Takeda Pharmaceutical Company Limited, and Tiziana Pharma, Ltd. Dr Mezquita reported performing consulting or an advisory role for Roche Diagnostics, Roche, and Takeda Pharmaceutical Company Limited; lectures and educational activities for Bristol-Myers Squibb and Tecnofarma SA; and receiving travel, accommodations, and/or expenses from Roche Diagnostics and Bristol-Myers Squibb. Dr Caramella reported receiving personal fees Merck & Co, Bristol-Myers Squibb, and Pfizer, Inc. No other disclosures were reported.

Additional Information: The following centers contributed data for the initial study by Ferrara et al,[13] but did not participate directly in the present study: Pneumology Department, Centre Hospitalier Toulon Sainte-Musse, Toulon, France; Nuclear Medicine Department, Centre Hospitalier Toulon Sainte-Musse, Toulon, France; Pneumology Department, Centre Hospitalier Universitaire de Toulouse, Universite Paul Sabatier, Toulouse, France; Thoracic Oncology Department, Hopital Bichat-Claude Bernard, Universite Paris-Diderot, Paris, France; Medical Oncology Department, Institute Bergonie, Bordeaux, France; Medical Oncology Department, Hopital Avicenne, Bobigny, France; Service des Maladies Respiratoires, Centre Hospitalier Universitaire de Bordeaux, Bordeaux, France; and Pneumology Department, Centre Hospitalier Universitaire de Besancon, Besancon, France.

.REFERENCES

1. Robert C, Schachter J, Long GV, et al; KEYNOTE-006 investigators. Pembrolizumab versus ipilimumab in advanced melanoma. *N Engl J Med*. 2015;372(26):2521-2532. doi:10.1056/NEJMoa1503093

2. Brahmer J, Reckamp KL, Baas P, et al. Nivolumab versus docetaxel in advanced squamous-cell non–small cell lung cancer. *N Engl J Med*. 2015;373(2):123-135. doi:10.1056/NEJMoa1504627

3. Motzer RJ, Escudier B, McDermott DF, et al; CheckMate 025 Investigators. Nivolumab versus everolimus in advanced renal-cell carcinoma. *N Engl J Med*. 2015;373(19):1803-1813. doi:10.1056/NEJMoa1510665

4. Ferris RL, Blumenschein G Jr, Fayette J, et al. Nivolumab for recurrent squamous-cell carcinoma

of the head and neck. *N Engl J Med*. 2016;375(19):1856-1867. doi:10.1056/NEJMoa1602252

5. Wolchok JD, Hoos A, O'Day S, et al. Guidelines for the evaluation of immune therapy activity in solid tumors: immune-related response criteria. *Clin Cancer Res*. 2009;15(23):7412-7420. doi:10.1158/1078-0432.CCR-09-1624

6. Champiat S, Dercle L, Ammari S, et al. Hyperprogressive disease is a new pattern of progression in cancer patients treated by anti-PD-1/PD-L1. *Clin Cancer Res*. 2017;23(8):1920-1928. doi:10.1158/1078-0432.CCR-16-1741

7. Gomez-Roca C, Koscielny S, Ribrag V, et al. Tumour growth rates and RECIST criteria in early drug development. *Eur J Cancer*. 2011;47(17):2512-2516. doi:10.1016/j.ejca.2011.06.012

8. Le Tourneau C, Servois V, Diéras V, Ollivier L, Tresca P, Paoletti X. Tumour growth kinetics assessment: added value to RECIST in cancer patients treated with molecularly targeted agents. *Br J Cancer*. 2012;106(5):854-857. doi:10.1038/bjc.2012.10

9. Kato S, Goodman A, Walavalkar V, Barkauskas DA, Sharabi A, Kurzrock R. Hyperprogressors after immunotherapy: analysis of genomic alterations associated with accelerated growth rate. *Clin Cancer Res*. 2017;23(15):4242-4250. doi:10.1158/1078-0432.CCR-16-3133

10. Fuentes-Antrás J, Provencio M, Díaz-Rubio E. Hyperprogression as a distinct outcome after immunotherapy. *Cancer Treat Rev*. 2018;70:16-21. doi:10.1016/j.ctrv.2018.07.006

11. Saâda-Bouzid E, Defaucheux C, Karabajakian A, et al. Hyperprogression during anti-PD-1/PD-L1 therapy in patients with recurrent and/or metastatic head and neck squamous cell carcinoma. *Ann Oncol*. 2017;28(7):1605-1611. doi:10.1093/annonc/mdx178

12. Singavi AK, Menon S, Kilari D, et al. 1140PDPredictive biomarkers for hyper-progression (HP) in response to immune checkpoint inhibitors (ICI): analysis of somatic alterations (SAs). *Ann Oncol*. 2017;28(5). doi:10.1093/annonc/mdx376.006

13. Ferrara R, Mezquita L, Texier M, et al. Hyperprogressive disease in patients with advanced non-small cell lung cancer treated with PD-1/PD-L1 inhibitors or with single-agent chemotherapy. *JAMA Oncol*. 2018;4(11):1543-1552. doi:10.1001/jamaoncol.2018.3676

14. Schwartz LH, Litière S, de Vries E, et al. RECIST 1.1-update and clarification: from the RECIST committee. *Eur J Cancer*. 2016;62:132-137. doi:10.1016/j.ejca.2016.03.081

15. Anderson JR, Cain KC, Gelber RD. Analysis of survival by tumor response. *J Clin Oncol*. 1983;1(11):710-719. doi:10.1200/JCO.1983.1.11.710

16. Kim CG, Kim KH, Pyo K-H, et al. Hyperprogressive disease during PD-1/PD-L1 blockade in patients with non-small-cell lung cancer. *Ann Oncol*. 2019;30(7):1104-1113. doi:10.1093/annonc/mdz123

European Radiology
https://doi.org/10.1007/s00330-020-06784-y

ONCOLOGY

Association of metastatic pattern and molecular status in stage IV non-small cell lung cancer adenocarcinoma

Alison Dormieux[1] · Laura Mezquita[2] · Paul Henry Cournede[3] · Jordi Remon[4] · Melodie Tazdait[1] · Ludovic Lacroix[5] · Etienne Rouleau[5] · Julien Adam[6] · Maria-Virginia Bluthgen[2] · Francesco Facchinetti[7] · Lambros Tselikas[1] · Frank Aboubakar[2] · Charles Naltet[2] · Pernelle Lavaud[2] · Anas Gazzah[2] · Cécile Le Pechoux[8] · Nathalie Lassau[1,9] · Corinne Balleyguier[1,9] · David Planchard[2] · Benjamin Besse[2] · Caroline Caramella[1,9,10]

Received: 26 August 2019 / Revised: 2 February 2020 / Accepted: 28 February 2020

Abstract

Objectives The aim of our study was to investigate the association between driver oncogene alterations and metastatic patterns on imaging assessment, in a large cohort of metastatic lung adenocarcinoma patients.

Methods From January 2010 to May 2017, 550 patients with stage IV lung adenocarcinoma with molecular analysis were studied retrospectively including 135 *EGFR*-mutated, 81 *ALK*-rearrangement, 47 *BRAF*-mutated, 141 *KRAS*-mutated, and 146 negative tumors for these 4 mutations (4N). After review of the complete imaging report by two radiologists (junior and senior) to identify metastatic sites, univariate correlation analyzes were performed.

Results We found differences in metastatic tropism depending on the molecular alteration type when compared with the non-mutated 4N group: in the *EGFR* group, pleural metastases were more frequent (32% versus 20%; $p = 0.021$), and adrenal and node metastases less common (6% versus 23%; $p < 0.001$ and 11% versus 23%; $p = 0.011$). In the *ALK* group, there were more brain and lung metastases (respectively 42% versus 29%; $p = 0.043$ and 37% versus 24%; $p = 0.037$). In the *BRAF* group, pleural and pericardial metastases were more common (respectively 47% versus 20%; $p < 0.001$ and 11% versus 3%; $p = 0.04$) and bone metastases were rarer (21% versus 42%; $p = 0.011$). Lymphangitis was more frequent in *EGFR*, *ALK*, and *BRAF* groups (respectively 6%, 7%, and 15% versus 1%); $p = 0.016$; $p = 0.009$; and $p < 0.001$.

Conclusion The application of these correlations between molecular status and metastatic tropism in clinical practice may lead to earlier and more accurate identification of patients for targeted therapy.

✉ Caroline Caramella
caroline.caramella@gustaveroussy.fr

[1] Imaging Department, Gustave Roussy, Université Paris-Saclay, F-94805 Villejuif, France

[2] Cancer Medicine Department, Gustave Roussy, Université Paris-Saclay, F-94805 Villejuif, France

[3] MICS laboratory, CentraleSupélec, Université Paris-Saclay, Gif-sur-Yvette, France

[4] Medical Oncology Department, Centro Integral Oncología Clara Campal Bacelona, HM-Delfos, Barcelona, Spain

[5] Molecular Biology Department, Gustave Roussy, Université Paris-Saclay, F-94805 Villejuif, France

[6] Pathology Department, Gustave Roussy, Université Paris-Saclay, F-94805 Villejuif, France

[7] Research Department (U981), Gustave Roussy Cancer Campus, Université Paris-Saclay, F-94805 Villejuif, France

[8] Radiation Therapy Department, Gustave Roussy, Université Paris-Saclay, F-94805 Villejuif, France

[9] IR4M, UMR 8081, CNRS, Université Paris-Saclay, F-91400 Orsay, France

[10] Radiology Department, Gustave Roussy, 114 Rue Édouard-Vaillant, 94805 Villejuif Cedex, France

Published online: 23 April 2020

Key Points

- *Bone and brain metastasis are the most common organs involved in lung adenocarcinoma but the relative incidence of each metastatic site depends on the molecular alteration.*
- *EGFR-mutated tumors preferentially spread to the pleura and less commonly to adrenals, ALK-rearrangement tumors usually spread to the brain and the lungs, whereas BRAF-mutated tumors are unlikely to spread to bones and have a serous (pericardial ad pleural) tropism.*
- *These correlations could help in the clinical management of patients with metastatic lung adenocarcinoma.*

Keywords Adenocarcinoma of lung · Neoplasm metastasis · Multimodal imaging · Mutation

Abbreviations

FISH	Fluorescent in situ hybridization
IHC	Immunohistochemistry
NSCLC	Non-small cell lung cancer
OS	Overall survival

Introduction

Lung cancer remains the first cause of cancer death worldwide, with an increased proportion in certain populations, such as female or non-smokers [1–3]. Most patients are still diagnosed in advanced stage, on which the systemic therapies are the main therapeutic strategy. However, the 5-year overall survival (OS) rate still remains inferior to 6% [4].

The discovery of oncogene driver alterations represents one of the most important progress in non-small cell lung cancer (NSCLC) patients, mainly somatic mutations (i.e., *EGFR*, *BRAF*, etc.) or chromosomal rearrangements (i.e., *ALK*, *ROS1*, etc.). These genomic alterations are responsible for more than 30% of NSCLC [5]. The development of targeted therapies against these specific molecular alterations has substantially improved the prognosis and quality of life for these populations [6, 7].

Certain clinical phenotypes are commonly associated with these molecular alterations, i.e., younger age, adenocarcinoma histology, or non-smoking habit [8–10], molecular testing is mandatory to prescribe targeted therapy, currently indicated at diagnosis of advanced disease in clinical routine, the detection of *EGFR*, *BRAF* mutations, and *ALK* and *ROS1* rearrangements [11].

Unfortunately, in up to 30% of the patients, the sample tissue is insufficient to perform molecular analyses after pathological assessment [12], and sometimes systemic therapy may be initiated even in the absence of known molecular status.

A better knowledge of the clinical phenotype of these patients can help to prioritize the tissue specimen for molecular testing, or can be a strong argument for performing a re-biopsy. Therefore, the radiological pattern could provide additional information. Small previous studies have been reported with no solid data (Table 1).

EGFR is the most frequently investigated mutation, but results are contradictory (lung, liver, and brain are more affected in a study of 456 *EGFR* patients [13], pleura and bone in another study of 218 *EGFR* patients [14]).

We aimed to investigate the association between metastatic patterns and driver oncogene alterations in a large cohort of metastatic lung adenocarcinoma patients.

Methods

Study population

Advanced metastatic NSCLC patients diagnosed between January 2010 and May 2017 in one tertiary oncological center were retrospectively assessed.

A complete imaging work-up (contrast-enhanced computed tomography included chest and upper abdomen, brain imaging, and body 18FluoroDG-PET) and the molecular analysis including at least *EGFR*, *KRAS*, *BRAF*, and *ALK* molecular alterations performed on the primary or metastatic tissue were mandatory for inclusion. Patients with previous medical of other metastatic cancer were excluded.

This study was approved by the institutional review board. No informed consent was required. Enrollment was conducted in two steps: during the first phase, between January 2010 and May 2015, enrolling all patients with metastatic lung adenocarcinoma, then, during the second phase, between May 2015 and May 2017 enrolling *EGFR*-mutated, *ALK*-rearranged, *or BRAF*-mutated patients.

Clinical data were retrospectively collected, including gender, age at inclusion, and smoking status, among others.

Molecular analysis

EGFR, *BRAF*, and *KRAS* mutations were analyzed by genome sequencing, according to the clinical routine in our institution. *ALK*-rearrangement was investigated either by fluorescent in situ hybridization (FISH) or by immunohistochemistry (IHC). We defined as control group a "fourth-negative" group (4N), in case of *EGFR*, *BRAF*, *KRAS*, and *ALK* negative detection by a clinical routine testing.

Table 1 Literature data on the association of metastatic organs and molecular status *X*/*Y*: *X* represents the number of patients carrying the mutation, and *Y* the control group not mutated or carrying another mutation when specified. Vs: versus

	pleura	lung	Liver	brain	bone	pericardium	adrenal gland	lymphangitis
More frequent	***EGFR*** -[16]: 139/88 -[17] : 71/720 -[18]: 138/144 -[14]: 218/1008 ***ALK*** -[31] : 41/80 -[35] : 31/254 -[32]Vs EGFR: 68/130	***EGFR*** -[22]: 98/148 -[27]: 103/160 -[28]: 116/166 -[29]: 22/33 -[30]Vs ALK-KRAS: 126/47-35 -[13]: 456/607 ***KRAS*** -[35]: 64/221 -[28]: 31/166 -[14]: 784/1008 ***HER2*** -[36]: 65/3735	***EGFR*** -[31]: 39/80 -[18] : 138/144 -[13] exon 21: 190/607 -[26]: 27/454 ***ALK*** -[31]: 41/80 -[14]: 42/1008	***EGFR*** -[19]exon 19: 18/31 -[20]: 37/63 -[21]: 12/19 -[22] : 98/148 -[23]: 138/176 -[24]: 16/79 -[25] : 108/126 -[13]: 456/607 -[26]: 27/454	***EGFR*** -[14]: 218/1008 ***KRAS*** -[26]: 91/390	***EGFR*** -[17]: 71/720 ***ALK*** -[31]: 41/80 -[32]Vs EGFR: 68/130		***ALK*** Versus EGFR -[30]: 47/126 -[32]: 68/130
Less frequent	***EGFR*** -[27]: 103/160 ***KRAS*** -[37]: 143/357 -[17]: 277/596 -[30]Vs ALK-EGFR: 35/47-126		***KRAS*** [37] : 143/357	***EGFR*** -[14] : 218/1008		***KRAS*** -[17] : 277/596	***EGFR*** -[14]: 218/1008	
No difference	***EGFR*** -[22]: 98/148 -[38]Vs ALK: 118/33	***EGFR*** -[31]: 39/80 -[18]: 138/144 ***ALK*** -[31]: 41/80	***EGFR*** -[22]: 98/148 -[38]Vs ALK: 118/44	***EGFR*** -[31] : 39/80 -[19] exon 21: 8/31 -[38]Vs ALK: 118/33 -[39] : 62/62; Vs KRAS: 62/65 ***ALK*** -[31]: 41/80 ***KRAS*** -[37]: 143/357 -[39] : 65/62; Vs EGFR: 65/62	***EGFR*** -[31]: 38/90 -[22]: 98/148 -[38]Vs ALK: 118/33 -[39]: 62/62; Vs KRAS: 62/65 ***ALK*** -[31]: 41/80 ***KRAS*** -[37]: 143/357 -[39]: 65/62; Vs KRAS: 65/62	***EGFR*** -[38]Vs ALK: 118/33	***EGFR*** -[31]: 38/80 -[22]: 98/148 ***ALK*** -[31]: 41/80 ***KRAS*** -[37]: 143/357	

Imaging assessment

The imaging review was performed by two independent radiologists A.D. (last year radiology fellow) and C.C. (senior lung cancer radiologist with 12 years of experience), blinded to the molecular status. Discrepancies were reviewed and consensus obtained. Metastasis sites were described and classified in different subgroups: bone, brain, lung, pleura, adrenal gland, lymph node, liver, peritoneum, carcinomatous lymphangitis, spleen, soft tissue, pericardium, skin, kidney, pancreas, and thyroid. Primary tumor of the lung and mediastinal lymph nodes was not considered as metastatic (according to the 8th edition of the TNM classification [15]).

Any histologically proven lesion or lesions whose appearance on imaging (conventional and metabolic) and evolution were consistent with the diagnosis were considered metastatic.

Statistical analysis

Univariate statistical analysis was performed with SPSS (version 20; SPSS Inc.). The rate of metastasis organ by organ was compared is each molecular subgroup (*EGFR*+, *BRAF*+, *KRAS*+, and *ALK*+), to the quadruple-negative subgroup (4N) using Fisher's exact test or chi-square test when applicable.

The age difference between molecular group and 4N group was compared by using Student's *t* test.

Clinical characteristics (including gender and smoking status) were described according to molecular status, and differences were assessed by the chi-square test.

A *p* value less than 0.05 was considered statistically significant.

Results

Study population

According to the inclusion and exclusion criteria, from the original database of 939 NSCLC patients, the final study cohort included 550 patients with stage IV lung adenocarcinoma (Fig. 1).

Among them, 444 patients were initially diagnosed with stage IV disease, and the remaining 106 were included at their first imaging showing progression to metastatic stage.

The study population consisted in 294 women (45%) and 362 men (55%), with a mean age of 59 years (23–88).

This cohort was composed of 135 *EGFR*-mutated, 81 *ALK*-rearranged, 47 *BRAF*-mutated, 141 *KRAS*-mutated, and 146 quadruple-negative patients (4N).

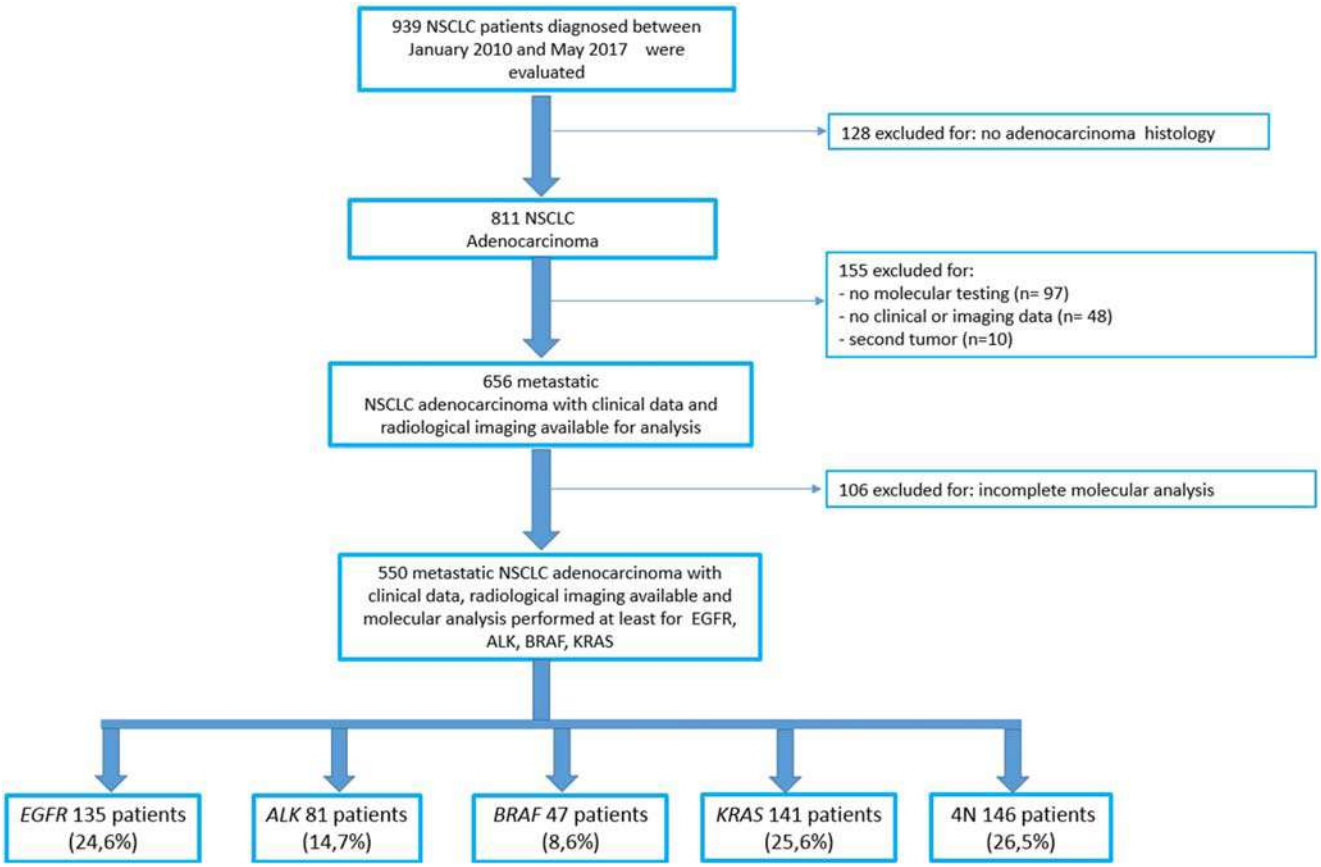


Fig. 1 Study flow chart

Table 2 Baseline characteristics of the study population

	EGFR $n = 135$ (%)	*ALK* $n = 81$ (%)	*BRAF* $n = 47$ (%)	*KRAS* $n = 141$ (%)	*4N* $n = 146$ (%)
Gender					
Female	96 (71)***	39 (48)	21 (45)	52 (37)	60 (41)
Male	39 (29)	42 (52)	26 (55)	89 (63)	86 (59)
Median, range	59, 25–88	52***, 23–82	63*, 35–88	59, 30–83	59, 29–89
Smoking status					
Non-smoker	78 (58)***	37 (46)***	17 (36)*	6 (4)***	28 (19)
Former smoker	40 (30)	26 (32)	16 (34)	55 (39)	48 (33)
Current smoker	17 (13)	18 (22)	14 (30)	80 (57)	70 (48)

*p value compared with control group 4N < 0.05

***< 0.001

Among the *EGFR*-mutated patients, 78 had *EGFR* exon 19 deletions, 40 *EGFR* L858R exon 21 mutations, 7 *EGFR* exon 18 mutations, and 10 *EGFR* exon 20 mutations.

Among the *KRAS*-mutated patients, mainly G12 ($n = 102$) (subtype C ($n = 59$) and V ($n = 25$)), G13 ($n = 11$), or uncommon ($n = 28$).

In case of *BRAF*-mutated patients, 40 had $BRAF^{V600E}$ mutation and 7 other uncommon mutations.

Compared with the 4N group, *EGFR*-mutated, *ALK*-rearranged, and *BRAF*-mutated patients were significantly associated with non-smoking status (respectively $p < 0.001$; $p < 0.001$; *and* $p < 0.05$); in contrast *KRAS*-mutated patients were more commonly smokers ($p < 0.001$).

ALK-rearranged and *BRAF*-mutated patients were the youngest and the oldest populations in the study, respectively ($p < 0.001$ *and* $p < 0.05$).

EGFR-mutated group had significantly more female than 4N group ($p < 0.001$) (Table 2).

Molecular alterations and metastatic pattern

Overall, the most common sites of metastatic disease were bone (44%), brain (34%), and lung (30%).

The distribution of metastatic organs according to the molecular alteration is described in Table 3 and Fig. 2.

Pleural metastases and lung lymphangitis were significantly more frequent in *EGFR*-mutated patients vs. 4N group (32% vs. 20%, respectively; $p = 0.021$ and 6% vs 1%; $p = 0.016$). Adrenal gland, lymph node, and soft tissue metastases were significantly less common in the EGFR group (respectively 6% versus 23%; $p < 0.001$; 11% versus 23%; $p = 0.011$; and 1% versus 8%; $p = 0.016$).

In *ALK*-rearranged patients, the involvement of central nervous system (42% vs. 29%; $p = 0.043$), lung (37% versus 24%; $p = 0.037$), and lymphangitis (7% versus 1%; $p = 0.009$) were more common, compared with 4N group.

Table 3 Distribution of metastatic sites according to the molecular alteration

	EGFR $n = 135$ (%)	*ALK* $n = 81$ (%)	*BRAF* $n = 47$ (%)	*KRAS* $n = 141$ (%)	*4N* $n = 146$ (%)
Bone	71 (53)	35 (43)	10 (21)*	66 (47)	61 (42)
Brain	48 (36)	34 (42)*	10 (21)	51 (36)	42 (29)
Lung	43 (32)	30 (37)*	14 (30)	41 (29)	35 (24)
Pleura	43 (32)*	19 (23%)	22 (47)***	23 (16)	29 (20)
Adrenal gland	8 (6)***	12 (15)	5 (11)	38 (27)	34 (23)
Lymph node	15 (11)*	17 (21)	7 (15)	20 (14)	33 (23)
Liver	26 (19)	15 (19)	6 (13)	24 (17)	20 (14)
Peritoneum	1 (1)	3 (4)	4 (9)	9 (6)	7 (5)
Lymphangitis	8 (6)*	6 (7)**	7 (15)***	3 (2)	1 (1)
Pericardium	2 (1)	3 (4)	5 (11)*	3 (2)	4 (3)
Soft tissue	2 (1)*	2 (2)	1 (2)	10 (7)	11 (8)

*p value compared with control group 4N < 0.05

**< 0.01

***< 0.001

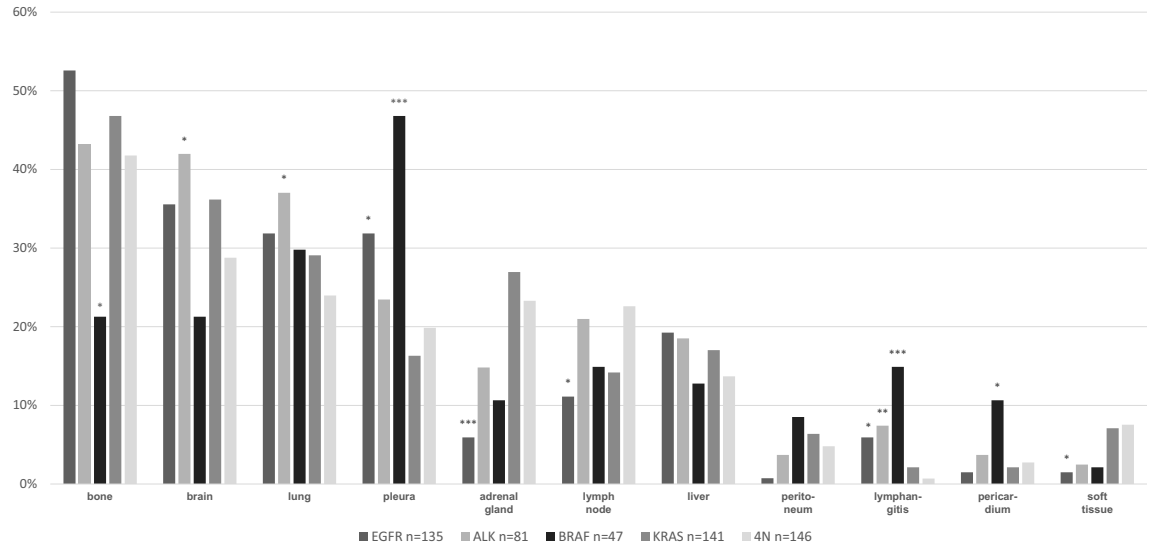


Fig. 2 Comparison of the distribution of metastasis sites by molecular subgroup. *p value compared with control group 4N < 0.05; **< 0.01; ***< 0.001

In *BRAF*-mutated patients, pleural (47% versus 20%; $p < 0.001$), pericardial metastases (11% versus 3%; $p = 0.040$), and lymphangitis (15% versus 1%; $p < 0.001$) were more frequent compared with 4N group; however, bone metastases were less common (21% vs. 42%; $p = 0.011$).

No differences were observed in the *KRAS*-mutated population compared with 4N group.

Discussion

We report in this article the correlation between molecular status and metastatic tropism in a large cohort of mutated and non-mutated lung adenocarcinoma patients. The strength of the present study relies on 2 main aspects: the retrospective reading of the complete multimodality imaging work-up by 2 dedicated radiologists and the number of molecular alterations investigated.

Similarly to the majority of previously published studies, *EGFR*-mutated patients have a higher rate of pleural metastases [14, 16–18] and a lower rate of adrenal metastases [14]. We found no difference on the incidence of brain metastases in the *EGFR* group compared with the 4N group according to previous studies (Table 1). Our results seem conflicting with those of Kuijpers et al who found a lower rate of brain involvement, but they considered neurologically asymptomatic patients without brain imaging as non-metastatic, which probably resulted in an under-diagnostic of this location [14]. On the contrary, some other authors described a higher rate of brain metastasis (interestingly, these are mainly Asian studies [13, 19–26]). A lower proportion of lymph node and soft tissue metastases were noted in the EGFR group, as well as a higher rate of lymphangitis: these observations have, to the best of our knowledge, never been described in the literature

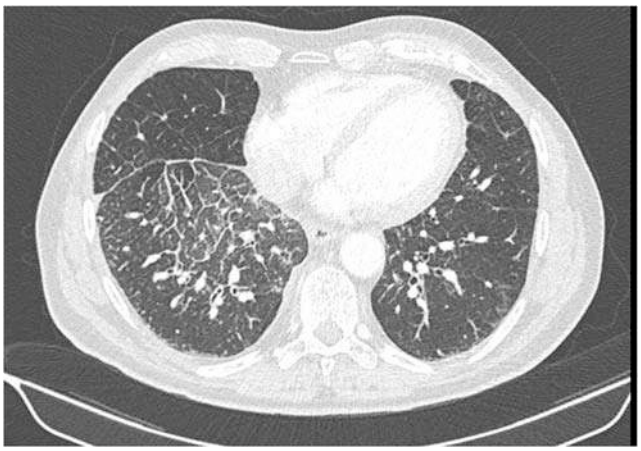


Fig. 3 Axial computed tomography (CT) image of a 72-year-old man with lung adenocarcinoma EGFR+, showing a carcinomatous lymphangitis of lower right lobe

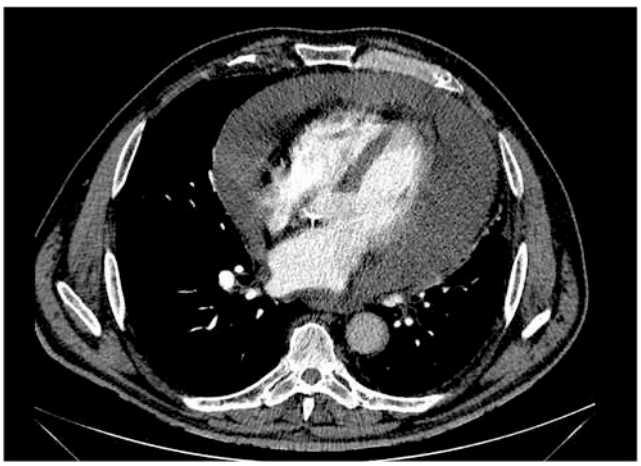


Fig. 4 Axial CT image of a 35-year-old woman patient with lung adenocarcinoma BRAF+, showing pericardial effusion

(Fig. 3). Several studies showed a higher incidence of lung metastases in *EGFR* group [13, 22, 27–30], but this trend does not appear significant in our study, similarly to the studies of Zhao et al and Doebele et al [18, 31].

In the *ALK* group, there were significantly more brain and lung metastases, which was not demonstrated in the Doebele et al's cohort of 41 patients [31].

Lymphagitis was more common in the *ALK* group, according to 2 other studies [30, 32].

Contrary to the literature, no significant difference was noted in regard to the rate of liver metastases [14, 31].

To our knowledge, our study is the first to describe *BRAF* lung tumors tropism as a unique pattern: pleural and pericardial metastases were significantly more frequent, confirming the findings of 3 previous case reports [33, 34] (Fig. 4). *BRAF*-mutated tumors also shown more lymphangitis and significantly inferior tropism for bone metastases compared with the 4N group.

In light of our results, we believe that the clinical management of patients with a diagnosis of metastatic lung adenocarcinoma could be transformed. In particular, It could improve and personalize the imaging interpretation by accurately identify metastases in the case of oligometastatic disease.

Our study has several limitations. First, there is a selection bias, related to the monocentric tertiary referral center effect. This could explain in particular the younger mean age of our patients than in other studies. The retrospective nature of our study is also a limitation; some patients had received systemic treatment prior to stage IV diagnosis, which may have changed the metastatic spread. There were a lot of excluded patients in the cohort, related to incomplete molecular profile, and we could not take into account other drivers such as *ROS1*, *RET*, *NTRK* fusions and *HER2* mutation, *MET* mutation, because of their rarity.

In conclusion, the application of these results in clinical practice will potentially lead to earlier and more accurate identification of patients for targeted therapy on basic imaging work-up.

Acknowledgments We thank Dr. Denoiseux for editing.

Funding information The authors state that this work has not received any funding.

Compliance with ethical standards

Guarantor The scientific guarantor of this publication is Dr. Caroline Caramella.

Conflict of interest The authors of this manuscript declare no relationships with any companies whose products or services may be related to the subject matter of the article.

Statistics and biometry One of the authors has significant statistical expertise.

Informed consent Written informed consent was not required for this study because of the retrospective design.

Ethical approval Institutional Review Board approval was obtained.

Methodology
- Prospective
- Observational
- Performed at one institution

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 68(6):394–424
2. INCA - Les cancers en France [Internet]. [cited 2019 Mar 27]. Available from: https://www.e-cancer.fr/ressources/cancers_en_france/
3. Locher C, Debieuvre D, Coëtmeur D et al (2013) Major changes in lung cancer over the last ten years in France: the KBP-CPHG studies. Lung Cancer 81(1):32–38
4. Cancer of the Lung and Bronchus - Cancer Stat Facts [Internet]. SEER. [cited 2019 Aug 22]. Available from: https://seer.cancer.gov/statfacts/html/lungb.html
5. Barlesi F, Mazieres J, Merlio J-P et al (2016) Routine molecular profiling of patients with advanced non-small-cell lung cancer: results of a 1-year nationwide programme of the French Cooperative Thoracic Intergroup (IFCT). Lancet 387(10026):1415–1426
6. Peters S, Camidge DR, Shaw AT et al (2017) Alectinib versus crizotinib in untreated ALK-positive non-small-cell lung cancer. N Engl J Med 377(9):829–838
7. Soria J-C, Ohe Y, Vansteenkiste J et al (2018) Osimertinib in untreated EGFR-mutated advanced non-small-cell lung cancer. N Engl J Med 378(2):113–125
8. Zou J, Lv T, Zhu S et al (2017) Computed tomography and clinical features associated with epidermal growth factor receptor mutation status in stage I/II lung adenocarcinoma. Thorac Cancer 8(3):260–270
9. Liu Y, Kim J, Qu F et al (2016) CT features associated with epidermal growth factor receptor mutation status in patients with lung adenocarcinoma. Radiology. 280(1):271–280
10. Solomon B, Varella-Garcia M, Camidge DR (2009) ALK gene rearrangements: a new therapeutic target in a molecularly defined subset of non-small cell lung cancer. J Thorac Oncol 4(12):1450–1454
11. Planchard D, Popat S, Kerr K et al (2018) Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol 29(Supplement_4):iv192–iv237
12. Zugazagoitia J, Ramos I, Trigo JM et al (2019) Clinical utility of plasma-based digital next-generation sequencing in patients with advance-stage lung adenocarcinomas with insufficient tumor samples for tissue genotyping. Ann Oncol 30(2):290–296
13. Li H, Cao J, Zhang X et al (2017) Correlation between status of epidermal growth factor receptor mutation and distant metastases of lung adenocarcinoma upon initial diagnosis based on 1063 patients in China. Clin Exp Metastasis 34(1):63–71
14. Kuijpers CCHJ, Hendriks LEL, Derks JL et al (2018) Association of molecular status and metastatic organs at diagnosis in patients

with stage IV non-squamous non-small cell lung cancer. Lung Cancer 121:76–81
15. Goldstraw P, Chansky K, Crowley J et al (2016) The IASLC lung cancer staging project: proposals for revision of the TNM stage groupings in the forthcoming (eighth) edition of the TNM classification for lung cancer. J Thorac Oncol 11(1):39–51
16. Wu S-G, Gow C-H, Yu C-J et al (2008) Frequent epidermal growth factor receptor gene mutations in malignant pleural effusion of lung adenocarcinoma. Eur Respir J 32(4):924–930
17. Smits AJJ, Kummer JA, Hinrichs JWJ et al (2012) EGFR and KRAS mutations in lung carcinomas in the Dutch population: increased EGFR mutation frequency in malignant pleural effusion of lung adenocarcinoma. Cell Oncol (Dordr) 35(3):189–196
18. Zhao J, Dinkel J, Warth A et al (2017) CT characteristics in pulmonary adenocarcinoma with epidermal growth factor receptor mutation. PLoS One 12(9):e0182741
19. Sekine A, Kato T, Hagiwara E et al (2012) Metastatic brain tumors from non-small cell lung cancer with EGFR mutations: distinguishing influence of exon 19 deletion on radiographic features. Lung Cancer 77(1):64–69
20. Ge M, Zhuang Y, Zhou X, Huang R, Liang X, Zhan Q (2017) High probability and frequency of EGFR mutations in non-small cell lung cancer with brain metastases. J Neurooncol 135(2):413–418
21. Matsumoto S, Takahashi K, Iwakawa R et al (2006) Frequent EGFR mutations in brain metastases of lung adenocarcinoma. Int J Cancer 119(6):1491–1494
22. Fujimoto D, Ueda H, Shimizu R et al (2014) Features and prognostic impact of distant metastasis in patients with stage IV lung adenocarcinoma harboring EGFR mutations: importance of bone metastasis. Clin Exp Metastasis. 31(5):543–551
23. Shin D-Y, Na II, Kim CH, Park S, Baek H, Yang SH (2014) EGFR mutation and brain metastasis in pulmonary adenocarcinomas. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 9(2):195–199
24. Tomasini P, Serdjebi C, Khobta N et al (2016) EGFR and KRAS mutations predict the incidence and outcome of brain metastases in non-small cell lung cancer. Int J Mol Sci 18:17(12)
25. Han G, Bi J, Tan W et al (2016) A retrospective analysis in patients with EGFR-mutant lung adenocarcinoma: is EGFR mutation associated with a higher incidence of brain metastasis? Oncotarget. 7(35):56998–57010
26. Renaud S, Seitlinger J, Falcoz P-E et al (2016) Specific *KRAS* amino acid substitutions and *EGFR* mutations predict site-specific recurrence and metastasis following non-small-cell lung cancer surgery. Br J Cancer 115(3):346–353
27. Hasegawa M, Sakai F, Ishikawa R (2016) CT features of epidermal growth factor receptor–mutated adenocarcinoma of the lung: comparison with nonmutated adenocarcinoma. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer 11(6):819–826
28. Lv J, Zhang H, Ma J et al (2018) Comparison of CT radiogenomic and clinical characteristics between EGFR and KRAS mutations in lung adenocarcinomas. Clin Radiol 73(6):590.e1–590.e8
29. Togashi Y, Masago K, Kubo T et al (2011) Association of diffuse, random pulmonary metastases, including miliary metastases, with epidermal growth factor receptor mutations in lung adenocarcinoma. Cancer. 117(4):819–825
30. Park J, Kobayashi Y, Urayama KY, Yamaura H, Yatabe Y, Hida T (2016) Imaging characteristics of driver mutations in EGFR, KRAS, and ALK among treatment-naïve patients with advanced lung adenocarcinoma. PLoS One 11(8):e0161081
31. Doebele RC, Lu X, Sumey C et al (2012) Oncogene status predicts patterns of metastatic spread in treatment-naive nonsmall cell lung cancer. Cancer. 118(18):4502–4511
32. Choi C-M, Kim MY, Hwang HJ, Lee JB, Kim WS (2015) Advanced adenocarcinoma of the lung: comparison of CT characteristics of patients with anaplastic lymphoma kinase gene rearrangement and those with epidermal growth factor receptor mutation. Radiology. 275(1):272–279
33. Mufti M, Ching S, Farjami S, Shahangian S, Sobnosky S (2018) A case series of two patients presenting with pericardial effusion as first manifestation of non-small cell lung cancer with BRAF mutation and expression of PD-L1. World J Oncol 9(2):56–61
34. Gautschi O, Pauli C, Strobel K et al (2012) A patient with BRAF V600E lung adenocarcinoma responding to vemurafenib. J Thorac Oncol 7(10):e23–e24
35. Rizzo S, Petrella F, Buscarino V et al (2016) CT Radiogenomic characterization of EGFR, K-RAS, and ALK mutations in non-small cell lung cancer. Eur Radiol 26(1):32–42
36. Mazières J, Peters S, Lepage B et al (2013) Lung cancer that harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives. J Clin Oncol Off J Am Soc Clin Oncol 31(16):1997–2003
37. Lohinai Z, Klikovits T, Moldvay J, Ostoros G, Raso E, Timar J et al (2017) KRAS-mutation incidence and prognostic value are metastatic site-specific in lung adenocarcinoma: poor prognosis in patients with KRAS mutation and bone metastasis. Sci Rep 7:39721
38. Miao Y, Zhu S, Li H et al (2017) Comparison of clinical and radiological characteristics between anaplastic lymphoma kinase rearrangement and epidermal growth factor receptor mutation in treatment naïve advanced lung adenocarcinoma. J Thorac Dis 9(10):3927–3937
39. Hendriks LEL, Smit EF, Vosse BAH et al (2014) EGFR mutated non-small cell lung cancer patients: more prone to development of bone and brain metastases? Lung Cancer 84(1):86–91

Résumé de la thèse vulgarisée

L'imagerie joue un rôle central dans le diagnostic de cancer et l'évaluation de l'efficacité des traitements. Il existe de multiples façon d'envisager la recherche en imagerie afin de permettre de mieux caractériser les tumeurs et d'aider à une meilleure prise en charge.

Mon travail de thèse s'est intéressé à 3 domaines. Tout d'abord la radiomique, qui utilise les images médicales comme des données numériques mais dont j'ai voulu évaluer la robustesse en quantifiant les possibles variations de mesures intra ou inter-scanners. Ensuite l'immunothérapie, un nouveau type de traitement prometteur sur la survie des patients atteints de cancers avancés, mais qui occasionne des réponses atypiques comme la pseudoprogression ou l'hyperprogression. Je me suis donc également intéressée à l'évaluation de la réponse chez ces patients. Enfin, la dernière partie de mon travail a visé à développer un algorithme d'intelligence artificielle capable de prédire la probabilité de trouver une anomalie génomique d'intérêt à partir de données cliniques et radiologiques simples, qui pourrait bénéficier aux patients pour lesquels une analyse du tissu tumoral n'est pas possible.

Résumé vulgarisé en anglais

Imaging plays a central role in the diagnosis of cancer and the evaluation of the effectiveness of treatments. There are multiple ways of looking at imaging research in order to better characterize tumors and help in better management.

My thesis work focused on 3 areas. First, radiomics, which uses medical images as digital data but whose robustness I wanted to evaluate by quantifying the possible variations of intra or inter-scanners measurements. Secondly, immunotherapy, a new type of treatment which is promising for the survival of patients with advanced cancers, but which causes atypical responses such as pseudoprogression or hyperprogression. I was therefore also interested in the evaluation of the response in these patients. Finally, the last part of my work aimed at developing an artificial intelligence algorithm capable of predicting the probability of finding a genomic anomaly of interest from simple clinical and radiological data, which could benefit patients for whom tumor tissue analysis is not possible.

université
PARIS-SACLAY

ÉCOLE DOCTORALE
Cancérologie : biologie - médecine - santé (CBMS)