

# 固形がんの治療効果判定のための 新ガイドライン (RECIST ガイドライン) 日本語訳 JCOG 版

SPECIAL ARTICLE

## New Guidelines to Evaluate the Response to Treatment in Solid Tumors

Patrick Therasse, Susan G. Arbuck, Elizabeth A. Eisenhauer, Jantien Wanders,  
Richard S. Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke,  
Allan T. van Oosterom, Michaele C. Christian, Steve G. Gwyther

*Journal of the National Cancer Institute*, 2000, Vol 92, No. 3, 205–216

### ABSTRACT

細胞傷害性の抗がん剤は、当該薬剤による腫瘍縮小量に基づく抗腫瘍活性の評価プロセスを経て開発がなされてきた。1970年代後期に、国際対癌連合(International Union Against Cancer; UICC)および世界保健機関(World Health Organization; WHO)により、腫瘍縮小効果の評価体系として明確な規準が提唱された。1994年には、その後の経験や知見に基づいてこれらの規準の再評価に取り組むために、臨床研究を実施しているいくつかの組織によるグループが結成された。数年にわたる徹底的な議論を経たのち、旧規準に代わる新しいガイドラインができあがった。この活動と並行して、参加研究グループの1つにより、それまでの二次元的なアプローチに代わり腫瘍病変を一次的に測定することで奏効率が得られる1つのモデルが開発された。この新たなコンセプトは、Response Evaluation Criteria in Solid Tumors(RECIST) Groupにより大規模な検証が行われ、新しいガイドラインに統合された。本稿では、効果判定のさまざまな目的を明確にするために、多少哲学的な背景にも言及する。本稿で提案する効果判定法は、存在するすべての病変を、測定対象となる標的的病変と、測定対象とならない非標的的病変とに分け、その両者の効果を複合して治療の総合効果を決定するというものである。腫瘍病変の評価法はガイドライン中に簡潔にまとめ、詳細は付録Iに記載した。効果判定に関するその他の要件はすべて必要に応じて議論され、再検討のうえ改正が施された。



Excerpta Medica

Affiliations of authors:

P. Therasse, J. Verweij, M. Van Glabbeke, A. T. van Oosterom, European Organization for Research and Treatment of Cancer, Brussels, Belgium; S. G. Arbuck, R. S. Kaplan, L. Rubinstein, M. C. Christian, National Cancer Institute, Bethesda, MD; E. A. Eisenhauer, National Cancer Institute of Canada Clinical Trials Group, Kingston, ON, Canada; J. Wanders, New Drug Development Office Oncology, Amsterdam, The Netherlands; S. G. Gwyther, East Surrey Healthcare National Health Service Trust, Redhill, U.K.

Correspondence to:

Patrick Therasse, M.D., European Organization for Research and Treatment of Cancer Data Center, Avenue Mounier 83/11, 1200 Brussels, Belgium (e-mail: pth@eortc.be).

“References” に続く “Note” を参照。

Original English Language Edition © 2000 Oxford University Press

Japanese Language Edition © 2002 Excerpta Medica, Japan/Elsevier Science K.K.

This publication has been made possible through Grants-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare, entitled “Study on Establishment of Clinical Trial System of Good Quality through Cooperative Group Mechanism(11S-4)”.

JCOG\* 運営委員会

(2001年10月現在)

西條長宏, 下山正徳, 福田治彦, 山本精一郎, 大橋靖雄, 新美三由紀, 児玉哲郎, 大倉久直, 江口研二, 飛内賢正, 島田安博, 渡辺 亨, 久保田馨, 国頭英夫, 佐野 武, 池田 恢, 石倉 聡, 手島昭樹, 早川和重, 加藤治文, 吉田茂昭, 大津 敦, 笹子 充, 安藤暢敏, 堀田知光, 高嶋成光, 吉川裕之

日本語訳担当: 下山正徳, 福田治彦

Japanese Translation ©Japan Clinical Oncology Group

\* Japan Clinical Oncology Group

日本臨床腫瘍研究グループ

厚生労働省がん研究助成金指定研究班

- 11 指-1 「高感受性悪性腫瘍に対する標準的治療確立のための多施設共同研究」  
主任研究者: 堀田知光 (東海大学医学部)
- 11 指-2 「呼吸器悪性腫瘍に対する標準的治療確立のための多施設共同研究」  
主任研究者: 西條長宏 (国立がんセンター中央病院)
- 11 指-3 「消化器悪性腫瘍に対する標準的治療確立のための多施設共同研究」  
主任研究者: 吉田茂昭 (国立がんセンター東病院)
- 11 指-4 「多施設共同研究の質の向上のための研究体制確立に関する研究」  
主任研究者: 福田治彦 (国立がんセンター研究所)

## A. はじめに

抗がん剤に対する腫瘍の客観的な縮小効果を定義する試みが開始されたのは1960年代初期のことであった<sup>1,2)</sup>。1970年代の中期から後期にかけて、客観的な腫瘍縮小効果(objective tumor response)の定義が広く普及し、用いられるようになった。がん治療の成績について一貫性のある報告を行うには、共通の言語が必要なことが明らかになったためである。

世界保健機関(World Health Organization; WHO)の定義は、1979年にWHOハンドブック<sup>3)</sup>として公表され、1981年にはMillerら<sup>4)</sup>により*Cancer*誌上にも公表されて、世界中の研究者がもっともよく使用する規準となった。しかし、WHO規準を用いるうちにいくつか問題が生じてきた：1) WHOにより定義された測定可能病変および「評価可能」病変の大きさの変化を効果判定に組み入れる方法が、研究グループ間でさまざまである、2) 記録される病変の大きさの下限や病変数もさまざまである、3) 進行(progressive disease; PD)の定義について、単一病変の変化で定義する研究グループもあれば、腫瘍の総量(全病変の測定値の和)の変化で定義する研究グループもある、4) 新技術(CT, MRI)の導入によって、三次元測定を効果判定にどのように組み入れるかについても多少の混乱が生じた。

これらの諸問題により、WHO規準に対する数多くの独自の変更や細則が作られ、そのため研究機関のあいだで効果判定規準に比較可能性がないという状況が生じた。これは、まさに当初WHOが規準を作成することで回避しようとした事態そのものであった。こうした状況の中で、いくつかの研究グループの代表者が中心となって、現行の効果の定義を再検討し、WHO規準の改訂版を策定する作業が動き出した。改訂に際しては、これまで議論や意見の不一致があった領域についても可能な限り検討することになった。

この作業を通じて多くの原則が確認された。

1) 腫瘍の退縮を起こさない機序で作用する可能性のある「新」治療法が開発されつつあるという事実はあるものの、固形がんの大きさの客観的な変化を記載し続けることは当然必要である。したがって、完全奏効(complete response; CR)、部分奏効(par-

tial response; PR)、安定(stable disease; SD)、進行(progressive disease; PD)というWHOハンドブック<sup>3)</sup>でもともと用いられていた4つのカテゴリーは、いかなるかたちにせよ改訂後も用いられるべきである。

- 2) 将来、新たな治療法によって良好な成績が得られたときに、現在使用されている治療法の成績と比較できるようにする必要があるので、たとえ測定規準が異なるうとも新旧のガイドライン間にPRの意味や概念上の大きな相違が生じないようにすべきである。
- 3) いくつかの研究施設は、腫瘍量の変化や退縮に先行する腫瘍代謝の変化を測定する技術を既に有するが、これらの技術はいまだ広く普及しているわけではなく、検証されていない点も多い。さらに、これまででも、効果判定規準が有用であるということは、測定が精密であることによるのではないと考えられてきた。とくに、PRの定義はもともと恣意的なものであり、腫瘍総量の「50%の減少」が個々の患者にとって本質的な意味があるわけではない。腫瘍量の測定の精密性を向上させることそのものが重要な目的とはみなされていなかったのである。むしろ、方法論の標準化と単純化が求められていた。とはいえ、本論文で提唱されるガイドラインは、治療ががん患者にもたらすであろう利益(benefit)の予測に関して、客観的な腫瘍縮小効果よりもさらに信頼性の高い新たな代替エンドポイントを生み出すための新技術の開発を妨げるものではない。
- 4) 現行のWHO規準では(おもに測定誤差により)誤って増悪と判定されやすいことに気付いたSouthwest Oncology Group(SWOG)などの研究グループは、腫瘍の増大の程度がもっと大きくなればPDと判定しない規準を既に採用している<sup>5)</sup>。今回の改訂WHO規準はこのことも考慮して変更が加えられている(付録IIを参照)。
- 5) 今回の新しい規準は、最近話題となっている問題のすべてを解決するものではないが、規準の見直しは今後も継続されるであろうし、とくに以下の事項については近い将来検討がなされるであろう。
  - ・細胞増殖抑制性薬剤(cytostatic agents)の第II相試験において、腫瘍の縮小に代わって適切な

検討を可能にするような抗腫瘍活性の指標

- ・ 血清腫瘍マーカーを用いた効果の定義およびその正当性を検証するために推奨される方法論
- ・ 特有の複雑性を示す特殊な腫瘍や解剖学的部位

## B. 背景

本ガイドラインは大規模な国際共同研究の成果である。1994年に、European Organization for Research and Treatment of Cancer (EORTC), National Cancer Institute (米国 NCI), および National Cancer Institute of Canada Clinical Trials Group (カナダ NCI 臨床試験グループ) が、固形がんの治療効果判定に用いる現行の規準の再検討を主目的とした専門委員会(付録 III を参照)を設立した。3年にわたり専門委員会内で定例会議および意見交換が行われたのち、WHO 規準改訂版の草案が作成され、広く配布された(付録 IV を参照)。寄せられたコメント(回収率 95%)をまとめ、専門委員会内で討議が行われ、適切なコメントに基づいて改訂第2版が発行された。この第2版文書は、未解決の問題を討議して解決するためのコンセンサス会議へ(所属組織の代表として)参加することが要請されていた外部審査者に再び配布された(1998年10月)。本コンセンサス会議への参加者名簿は付録 IV に示したが、学界、産業界、規制当局の代表者が含まれている。コンセンサス会議で討議された勧告に基づいて第3版文書が作成され、学界で公開[American Society of Clinical Oncology(米国臨床腫瘍学会), 1999]されたのち、公式の出版に向けて1999年6月に *Journal of the National Cancer Institute* 誌に投稿された。

腫瘍縮小効果(tumor response)の評価が行われた4,000例以上の患者を含む共同研究のデータから、従来の二次元の測定値と二方向積の和を用いる方法に代わり、一次元測定および最長径の和を用いて効果判定の単純化を行うことが妥当であることが示されている。種々のデータベースを用いて行われた個々のレトロスペクティブ解析(2つのアプローチを比較する)の成績を付録 V に記載する。次章のガイドラインで示すこの新たなアプローチは、James ら<sup>6)</sup>によって提唱されたモデルに基づくものである。

## C. 固形がんの効果判定規準 (RECIST) ガイドライン

### 1. 序論

序論では、腫瘍縮小効果判定規準の定義、前提事項、目的について考察する。以下に示すガイドラインに従うことによって、臨床試験の結果がより統一的なかたちで報告されるようになる。ここでは治験薬単剤の試験について記述するが、原則は多剤併用療法、治験薬以外を含む治療、あるいは薬剤を含まない治療の試験においても同様である。

抗がん剤投与に伴う腫瘍縮小効果の評価は、概念的に明確に区別される少なくとも3つの重要な目的をもって行われる。

- ・ 早期臨床試験におけるプロスペクティブなエンドポイントとしての腫瘍縮小効果：

この設定では、客観的な腫瘍縮小効果は、薬剤あるいはレジメンが開発研究を続けるに値する有望な結果を示すかどうかの判断に用いられる。このような試験は、典型的には治験薬あるいは試験的レジメンに関する第II相試験(1.2.項を参照)であり、まさにこの状況で用いるために本ガイドラインが作成されたといえる。

- ・ ある特定の患者コホートに対する利益の評価を目的にデザインされた、より検証的な臨床試験におけるプロスペクティブなエンドポイントとしての腫瘍縮小効果：

このような試験としては、ランダム化比較試験あるいは併用薬物療法に関して歴史的対照(historical control)との比較を行う単アーム試験が多い。この設定では、客観的な腫瘍縮小効果は、イベント(死亡あるいは増悪)発生までの期間(time to event)や症状のコントロール(1.3.項を参照)などの別の臨床的な利益の指標に対する代替エンドポイント(surrogate end point)として使用される。

- ・ 施行中の治療を継続するかどうかを臨床医と患者/被験者が決定するための指標としての腫瘍縮小効果：

これは臨床試験および日常診療の双方の使用目的に適用できる(1.1.項を参照)が、治療継続の決定を目的とした使用は本ガイドラインの主旨ではない。

しかし、日常的に「腫瘍縮小効果(tumor response)」という用語を使用する場合には、これらの目的を意図的に区別する努力を払わない限り容易に混同が生じる。この違いが無視されると、不適切な方法論が用いられ、誤った結論が導かれる可能性がある。

### 1.1. がんの日常診療における治療効果の指標

がんの日常診療における腫瘍縮小効果の評価は、事前に決められた規準に則って行われるわけではない。むしろ、患者に対する治療利益の評価に用いられる臨床所見や検査データから導き出される主観的な医学的判断に基づいている。このような状況に完全に適応させるために、本ガイドラインの規準をさらに検討したり明確化したりする必要はない。臨床試験ではない日常の患者管理においては、「臨床的な改善」と「客観的な腫瘍縮小効果」は区別するほうが適切であろう。

### 1.2. 新しい治療法の開発研究を進める指針としての無対照試験における治療効果の指標

「観察された奏効率(observed response rate)」は、さらなる開発研究に値する新規抗がん剤の「スクリーニング(screen)」としての単アーム試験で用いられることが多い。関係するエンドポイントとして、奏効期間(response duration)やCR患者の割合などが同様の状況で用いられることがある。奏効率をこのように用いる場合、「効果」は治療による利益を必ずしも意味するわけではなく、むしろ試験薬剤がある程度の生物学的な抗腫瘍活性を有することを意味するのである。

あるタイプの薬剤(細胞傷害性薬剤, ホルモン剤など)では、自然経過をみた場合に予想される奏効率よりも高い奏効率を示した抗がん剤をさらなる研究のために選択するという方法が有用であることが経験的に知られている。この方法で選ばれた薬剤には最終的に臨床的な有用性が証明されたものもある。さらに、この方法で新規薬剤をスクリーニングするための規準は、蓄積された経験に基づいた改変が可能であるが、最終的にはその規準でスクリーニングされた薬剤のうちどれだけが、後続のより検証的な試験によって実際に臨床的価値が証明されたかという効率(efficiency)の観点からその妥当性が確認される必要がある。

しかし、開発研究を継続するに足るかどうかを判断するために前もって決められた奏効率を達成した新規抗がん剤でも、引き続いて実施されたランダム化第III相試験で研究対象の疾患に有効な治療法であると証明されないことも多い。ランダムなバラツキや選択バイアスは、それが既知の場合も未知の場合も、小規模の無対照試験では圧倒的な影響を及ぼす可能性がある。このような試験は、ある対象疾患に対する新規薬剤あるいは併用薬物療法の抗腫瘍活性の初期評価においては効率的で経済的なステップである。しかし、同様の試験が数多く実施されると、偽陽性(false-positive)がもたらされる可能性が必然的に生じる。多くの場合、大規模で多くのリソースを要する第III相試験を実施する前に、もう1つの小規模の確認的な試験を実施することが適切といえよう。

新たな治療方法がランダム化第II相試験で検討されることがある。第II相試験でランダム化を行う目的は、第III相試験と同様、予後因子の不均衡による影響を最小限にすることである。しかし、ランダム化第II相試験とはその定義自体からも、適切な検出力をもってアーム(レジメン)間の比較を行うことを意図するものではない。むしろその目的は、さらなる研究に値する1つもしくはそれ以上の治療アームを選択することに過ぎず、サンプルサイズは、真に劣ったアームが選択されにくくなるのに十分な信頼性をもつように決められる。したがって、このようなランダム化第II相試験の成績の報告には、治療アーム間の統計学的な比較を含めるべきではない。

### 1.3. 臨床試験における緩和効果の代替としての治療効果の指標

1.3.1. 非ランダム化試験での使用：非ランダム化試験における腫瘍縮小効果が(純粋に臨床的な利益の指標とはみなせない場合においても)緩和効果の指標であるとの仮定が許容される唯一の状況とは、既存の類似症例の集団との比較が、実際以上にせよ想定以上にせよ可能である場合である。この仮定がもっとも妥当となるのは、試験対象患者と、定義に従って集積された対照患者とのあいだで、適切な予後因子についてマッチングするために予め計画された統計手法が用いられる場合である。この条件が満たされない場合は、最低でも、臨床的状況に応じて未治療または既存の標準治療を受

けた適切な「対照」集団において期待される奏効率の仮定に対して、きわめて強力な正当性を付与する予め決められた統計学的な規準が存在する必要がある。しかし、このような状況下においてさえ、試験対象患者で観察された奏効率が高かったとしても、治療の臨床的な利益の証明にはならない。非ランダム化試験には回避不能で定量化できないバイアスが必ず含まれるため、利益の証明には十分な症例数を有するプロスペクティブなランダム化比較試験による最終的な検証が必要である。そのような検証的な試験における治療の利益に関する適切なエンドポイントは、生存期間(survival)、無増悪生存期間(progression-free survival)あるいは症状のコントロール[生活の質(QOL)を含む]である。

1.3.2. ランダム化試験での使用：プロスペクティブなランダム化第 III 相比較試験においても、「観察された奏効率」を単独の、または主要なエンドポイントとすべきではない。治療の利益を反映する、より確定的なエンドポイント(たとえば生存期間、無増悪生存期間、症状の軽減、QOL の改善もしくは維持)との関連を用いて奏効率の差を検証するには、試験の規模が十分に大きくなくてはならない。

## 2. ベースライン(治療前)における腫瘍病変の測定可能性

### 2.1. 定義

ベースラインにおいて、腫瘍病変は測定可能(measurable)か測定不能(nonmeasurable)のいずれかに分類される。測定可能病変とは、少なくとも一次元で正確に測定でき(最長径が記録されること)、従来の検査法で 20 mm あるいはヘリカル CT で 10 mm (2.2.項を参照)の病変であり、測定不能病変とはそれ以外のすべての病変であり、小病変(最長径が従来の検査法で <20 mm またはヘリカル CT で <10 mm)と真の測定不能病変を含む。

測定可能性に関して用いられてきた「評価可能な(evaluable)」という用語は、それを使うことによって規準の価値や正確性が高まるわけではないため推奨されず、使用しないこととする。

あらゆる測定値は、定規あるいはキャリパス(ノギス)を用いてメートル法にて記録する。ベースラインでの

すべての評価は、可能な限り治療開始と近接した時期に実施し、治療開始前 4 週間以内でなくてはならない。

真に測定不能とみなされる病変には以下のものがある。

骨病変、軟膜病変、腹水、胸水/心膜液、炎症性乳がん、皮膚リンパ管炎/肺リンパ管炎、画像検査で確認されない腹部腫瘍、嚢胞性病変である。

(注：放射線照射の既往がある部位に存在する腫瘍病変は、測定可能とみなされる場合もそうでない場合もありうる。こうした病変の取り扱いについては必要に応じてプロトコルにて定義すべきである。)

### 2.2. 各測定法のスペック(仕様)

ベースラインおよびフォローアップ期間中を通じて、個々の病変の評価、記録、報告には、同じ評価法、同じ技術を用いなければならない。ある治療法の抗腫瘍効果を判定するために、画像に基づく評価と臨床的な評価(診察による測定)の両方が用いられた場合は、画像による評価を採用するのが望ましい。

2.2.1. 臨床的な評価(診察による測定): 臨床的に同定された病変が測定可能とみなされるのは表在性(皮膚小結節、触知可能なリンパ節など)の場合に限られる。皮膚病変の場合は、病変の大きさが評価できるように定規を写し込んだカラー写真による記録が推奨される。

2.2.2. 胸部 X 線写真: 胸部 X 線で描出された病変は、その輪郭が明確に確認され肺野(空気を含む肺)で囲まれていれば測定可能病変としてよいが、CT がより好ましい。付録 I に、客観的な腫瘍縮小効果の判定に関するこの評価法の詳細を記載する。

2.2.3. CT, MRI: CT および MRI は、効果を評価するために選択した標的病変を測定するのに、現在もっとも広く利用可能で、かつもっとも再現性に優れた方法である。従来型の CT および MRI は、スライス厚 10 mm 以下の連続断面で撮像する。ヘリカル CT は 5 mm の連続再構成アルゴリズムを用いて撮像する。このスペックは胸部、腹部、骨盤内の腫瘍に適用されるが、頭頸部腫瘍および四肢の腫瘍には通常特殊なプロトコルが必要である。付録 I に、客観的な腫瘍縮小効果の判定に関するこの評価法の詳細を記載する。

2.2.4. 超音波: 試験のプライマリーエンドポイントが

客観的な腫瘍縮小効果である場合は、深部の腫瘍病変の測定に超音波を用いるべきではない。表在性の触知可能なリンパ節、皮下病変、および甲状腺腫瘍の臨床的測定(診察による測定)の代替法としてなら使用してもよい。臨床的に評価することが一般的である表在性病変の完全消失を確定する際にも超音波は有用である。客観的な効果判定において、腫瘍病変の測定に超音波を用いないことが正当化される理由については付録Iに記載した。

2.2.5. 内視鏡, 腹腔鏡: これらの技術の客観的な腫瘍評価における有用性は、いまだ十分に広範な規模では検証されていない。特別な状況でこれらを用いるには、一部の施設でのみ使用が可能な精密機器と高度な専門知識が求められる。したがって、この種の技術を客観的な腫瘍縮小効果の評価に使用する場合は、専門施設での評価の妥当性の確認という目的に限定すべきである。しかし、生検材料が採取できる場合の病理組織学的CRの確認には有用である。

2.2.6. 腫瘍マーカー: 腫瘍マーカーは、単独では効果の評価には使用できない。しかし、ベースラインでマーカーが正常上限を超えていた患者で、全病変が消失した場合、CRと判定されるには、マーカーが正常化している必要がある。前立腺特異抗原(PSA)およびCA125については、これらの変化を臨床試験での効果判定に標準的に使えるようにするための特別な付加的規準が作られ、現在その妥当性が検討されている。

2.2.7. 細胞診, 組織診: 細胞診および組織診は、特殊な場合にのみ、PRとCRの区別に用いられる(たとえば、胚細胞腫瘍のようながん種において、治療後に良性の遺残病変と悪性の遺残病変を鑑別する場合など)。また、測定可能病変が奏効(CR, PR)あるいはSDの規準を満たしている場合に、治療中に出現あるいは増悪した体腔液ががん性かどうかを確かめる目的で細胞診による確認が必要となる。このような状況では、奏効(CR, PR)もしくはSD(体腔液は治療の副作用の可能性もある)とPD(がん性であることが確認される場合)を鑑別するために、採取された体腔液の細胞診を利用してもよい。

客観的な腫瘍縮小効果をより適切に判定するための新技術は、腫瘍縮小効果の判定にそれらを用いることの妥当性が十分に確立されれば、将来この規準に組み込まれるだろう。

### 3. 腫瘍縮小効果の判定

#### 3.1. ベースライン評価

3.1.1. 全腫瘍量および測定可能病変の評価: 客観的な効果を評価するには、経過中の測定値と比較するためにベースラインにおける全腫瘍量(overall tumor burden)の評価が必要である。客観的な腫瘍縮小効果がプライマリーエンドポイントであるプロトコールでは、ベースラインにおいて測定可能病変を有する患者に対象を限定する。「測定可能な疾患(measurable disease)」とは、少なくとも1つの測定可能病変を有することである(2.1.項で定義)。測定可能病変が孤立性病変1つだけの場合は、その病変が当該がん種であることを細胞診または組織診にて確認する。

3.1.2. 「標的」および「非標的」病変のベースライン評価と記録: すべての測定可能病変のうち、すべての浸潤臓器の代表として1臓器につき最大5カ所、合計10病変までを標的病変として選択し、ベースラインにおいて測定、記録する。標的病変は大きさ(最長径を用いる)と、繰り返して正確に測定することに適しているかどうか(画像診断あるいは臨床的評価のいずれか)に基づいて選択する。すべての標的病変の最長径の和を算出し、ベースライン最長径和(baseline sum longest diameter)として記録する。ベースライン最長径和は、客観的な腫瘍縮小効果を評価するための比較対照とする。

標的病変以外のすべての病変(あるいは浸潤部位; site of disease)は非標的病変とし、ベースラインにおいて記録する。フォローアップ期間を通じて非標的病変の測定は必要ではないが、おのおのの病変の有無は記録しなければならない。

#### 3.2. 効果判定規準

3.2.1. 標的病変の評価: ここでは、標的病変における客観的な腫瘍縮小効果の判定に使用する規準の定義を示す。この規準では、最長径の測定が必要なのは標的病変のみであることを考慮しつつ、原本のWHOハンドブック<sup>3)</sup>の規準を適用した。

完全奏効(complete response; CR): すべての標的病変の消失。

部分奏効(partial response; PR): ベースライン最長径和と比較して標的病変の最長径の和が30%以上減少。

進行( progressive disease; PD ): 治療開始以降に記録された最小の最長径の和と比較して標的病変の最長径の和が 20% 以上増加。

安定( stable disease; SD ): PR とするには腫瘍の縮小が不十分で、かつ PD とするには治療開始以降の最小の最長径の和に比して腫瘍の増大が不十分。

3.2.2. 非標的病変の評価：ここでは、非標的病変における客観的な腫瘍縮小効果の判定に使用する規準の定義を示す。

完全奏効( complete response; CR ): すべての非標的病変の消失かつ腫瘍マーカー値の正常化。

不完全奏効/安定( incomplete response/stable disease; IR/SD ): 1 つ以上の非標的病変の残存かつ/または腫瘍マーカーが正常上限値を超える。

進行( progressive disease; PD ): 既存の非標的病変の明らかな増悪<sup>1)</sup>。

[ 注：「非標的」病変のみが明らかな増悪を示すことはまれであるが、そのような状況では、治療している医師の意見が尊重されるべきである。ただし、その増悪の状態は後日判定委員会( review panel )( または研究代表者 )により確定されるべきである。 ]

3.2.3. 最良総合効果の評価：最良総合効果( best overall response )とは、治療開始から増悪/再発までに記録された最良の効果のことである( PD では治療開始以降に記録された最小の測定値を評価の比較対照とする )。原則として、最良総合効果の判定は、測定値の規準と確定の規準( 3.3.1.項を参照 )の双方を用いて行う。表 1 は、標的病変と非標的病変および新病変の出現の有無を含めた腫瘍縮小効果の総合効果( overall response )の決定方法を示す。

[ 注：

- ・健康状態の全般的な悪化により、増悪( disease pro-

gression )の客観的証拠が得られないまま治療を中止せざるを得ない場合は「病状悪化( symptomatic deterioration )」に分類する。ただし、治療中止後も、客観的な増悪の証拠を得るようあらゆる努力を払うこと。

- ・早期増悪、早期死亡、評価不能と定義される可能性のある状態は各試験に特異的であるため、個々のプロトコールで明確に定義する( 治療期間や治療のコース間隔に依存する )。
- ・状況によっては、残存病変を正常組織と区別することが困難な場合がある。CR の評価がこの区別に依存する場合は、CR と確定する前に残存病変の検査( 細針吸引/生検 )を行うことが推奨される。]

3.2.4. 腫瘍再評価の頻度：治療中の腫瘍再評価の頻度はプロトコールごとに決めるべきものであり、治療のタイプおよびスケジュールに合わせて決定する。ただし、治療の有益な効果が未知である薬剤の第 II 相試験では、1 コースおき( 2 コースごと )のフォローアップ( 6 ~ 8 週ごと )は妥当な一般標準と考えられる。特殊なレジメンや状況によっては、これよりも短い間隔あるいは長い間隔が正当化される場合もあり得る。

治療終了後に腫瘍の評価を繰り返し行う必要があるかどうかは、その第 II 相試験の目的が奏効率の評価か、イベント( 増悪、死亡 )発生までの期間( time to an event )の評価であるかによる。試験の主要なエンドポイントがイベント発生までの期間である場合は、予期されたイベント以外の原因で治療中止となった患者についても、ルーチンの再評価の頻度をプロトコールで規定しておく。治療中止後の評価の間隔は治療中の評価間隔の 2 倍とすることが多いが、厳密な規則はない。

表 1 新病変出現の有無を含む標的病変と非標的病変の腫瘍縮小効果の組み合わせによる総合効果\*

標的病変	非標的病変	新病変	総合効果
CR	CR	なし	CR
CR	IR/SD	なし	PR
PR	PD 以外	なし	PR
SD	PD 以外	なし	SD
PD	いずれでもよい	いずれでもよい	PD
いずれでもよい	PD	いずれでもよい	PD
いずれでもよい	いずれでもよい	あり	PD

\* CR = complete response( 完全奏効 ), IR = incomplete response( 不完全奏効 ), PR = partial response( 部分奏効 ), SD = stable disease( 安定 ), PD = progressive disease( 進行 )。詳細は本文を参照。



### 3.3. 確定のための測定/奏効期間

3.3.1. 確定(confirmation): 臨床試験における客観的な効果を確定すること(confirmation)のおもな目的は、観察された奏効率の過大評価を回避することである。効果判定に関するこの考え方は、腫瘍縮小効果をプライマリーエンドポイントとする非ランダム化試験においてとくに重要である。こうした試験では、PRまたはCRと判定するには、最初にその効果の規準を満たしたときから4週以降に行われる再評価によって、その規準を再び満たすことを確認する必要がある。試験によってはプロトコルで、さらに長い間隔を規定することが適切である場合もある。

SDの場合は、プロトコルで定義する、試験登録後の最短の期間(一般に、6~8週以上)を越えて、少なくとも1度は測定値がSD規準を満たしていなければならない(3.3.3.項参照)。

[注: 無増悪生存期間(progression-free survival)や全生存期間(overall survival)を主要なエンドポイントとする試験においては、腫瘍縮小効果の確定のための反復検査は常に実施可能性があるとはいえず、標準的な方法とする必要はない。このような試験では「確定効果(confirmed response)」は用いない。こうした試験の結果を報告する際にはこの区別を明確にすべきである。]

3.3.2. 全奏効期間: 全奏効期間(duration of overall response)とは、CRあるいはPRの測定規準を満たした時点(最初に記録された状態がどちらであれ)から、再発あるいはPDが客観的に確認された最初の日(治療開始以降に記録された最小測定値をPDの評価の比較対照とする)までの期間と定義される。完全奏効期間(duration of overall complete response)は、はじめてCRの測定規準を満たした時点から、再発が客観的に確認された最初の日までの期間と定義される。

3.3.3. 安定期間: 安定期間(duration of stable disease)は、治療開始からPD規準が満たされるまでの期間と定義される(治療開始以降に記録された最小測定値を評価の比較対照とする)。安定期間の臨床的妥当性はがん種や悪性度(grade)の違いに応じて異なる。したがって、プロトコルにはSDを決める2回の測定に必要な最短の間隔を明記することが強く推奨される。この測定日の間隔は、SDであることが試験の対象集

団にもたらす予期される臨床的な利益を考慮して決める。

[注: 奏効期間や安定期間だけでなく、無増悪生存期間もベースライン評価後のフォローアップの頻度によって影響を受ける。標準的なフォローアップの頻度を定義するには、病型や病期、治療周期、標準的な日常診療などの多くのパラメーターを考慮する必要があるが、この点は本ガイドラインの範疇ではない。しかし、試験間の比較がなされる場合は、これら測定されたエンドポイントの精度(precision)の限界を考慮しなければならない。]

### 3.4. 無増悪生存期間/無増悪期間

ここでは、客観的な効果に関連するエンドポイントを中心に述べる。場合によっては[たとえば脳腫瘍や、腫瘍縮小効果を示さない抗がん剤(non cytoreductive anticancer agents)の研究など]、新規薬剤や新規レジメンの抗腫瘍活性を評価するためには、腫瘍縮小効果の判定は最適な方法ではないことがある。このような場合、非細胞傷害性の機序によって作用する新規薬剤の生物学的効果に関する初期評価を行うには、無増悪生存期間(progression-free survival)無増悪期間(time to progression)が考慮に値する代替指標となる可能性がある。しかし、無増悪生存期間/無増悪期間を用いる無対照試験では、治療効果がない場合に無増悪生存期間/無増悪期間がどの程度になることが推定されるかについて、注意深い記述が必要である。患者選択や増悪の確認などに関する交絡バイアスが生じやすい状況であると考えられる場合は、解析を非常に保守的(conservative)にすることも推奨される。無増悪生存期間/無増悪期間をプライマリーエンドポイントとして用いる無対照試験では、試験ごとに十分に検討し、適用する方法論をプロトコルにもれなく記載する必要がある。

## 4. 効果の審査(response review)

奏効率をプライマリーエンドポイントとする試験では、試験終了時に1名あるいは数名の、研究から独立した専門家により、すべての効果について審査を受けることを強く推奨する。患者ファイルとX線画像の同時審査が最良の方法である。

(注：放射線画像の審査を行う場合は、病変を不明瞭にしたり、審査者の評価にバイアスを生じさせるような印や書き込みが、画像上に残らないようにすることも推奨される。)

## 5. 結果の報告

プロトコル治療からの大きな逸脱 (major deviations) があつたり、不適格である場合も、試験に組み入れられたすべての患者の治療効果を評価しなければならない。個々の患者の治療効果は以下のカテゴリーのどれか1つに分類する。1) CR; complete response, 2) PR; partial response, 3) SD; stable disease, 4) PD; progressive disease, 5) 悪性疾患による早期死亡, 6) 毒性による早期死亡, 7) 他の原因による早期死亡, 9) 不明(評価不能, データ不十分)。[注：慣習的に、カテゴリー「9」は一般に、臨床データベース中のあらゆるタイプのデータに関して「不明な」状態を意味するコードとして使われる。8)の誤りではない。]

適格規準を満たしたすべての患者は、奏効率に関する主たる解析に含まなければならない。効果のカテゴリー4～9の患者は治療に奏効しなかったもの(増悪)とみなすべきである。したがって、治療スケジュールや薬剤投与が不正確だからといって奏効率の解析から除外してはならない。カテゴリー4～9の細かい定義はプロトコルごとに記述する。

すべての結論はすべての適格例の解析に基づいて下される必要がある。95%信頼区間を必ず付ける。

大きな逸脱例(たとえば、他の原因による早期死亡、早期の治療中止、大きなプロトコル違反など)を除いた患者のサブセットに基づくサブグループ解析が行われることがある。しかし、こうしたサブグループ解析は、治療効果に関する結論を出すための根拠を示すものとはみなされない。また、患者を解析から除外する理由は明確に報告しなければならない。

## 6. ランダム化第III相試験における効果判定

第III相試験における効果判定は、評価された治療の相対的な抗腫瘍活性の指標となる可能性があるが、一般的には試験の対象集団に対する実際の治療の利益を予測するものではない。客観的な腫瘍縮小効果を第

III相試験のプライマリーエンドポイントとして選択する場合(客観的な腫瘍縮小効果と実際の治療の利益との直接的な関係を、試験の対象集団において明瞭に示すことが可能な場合に限られる。)は、第II相試験で用いられる規準と同じ規準(RECISTガイドライン)を用いるべきである。

一方、客観的な腫瘍縮小効果をプライマリーエンドポイントとしない試験(たとえば第III相試験)では、本稿のガイドラインが必要とされない場合がある。たとえば、そのような試験では、10個もの標的病変を測定する必要はなく、4週以上の間隔をあけたフォローアップ評価で効果を確定する必要もない。プロトコルには、実際に用いる効果判定方法を明記し、また患者の評価可能性に影響を及ぼす後知恵的な(post-hoc)判定を避けるために、確定(confirmation)が必要であるか否かを明確に記述するべきである。

## 付録I. 放射線画像診断に関するスペック(仕様)

この注記事項は本ガイドラインを臨床研究で用いる際の推奨事項であり、ここで規定するCTやMRIのスキャンングの方法は、おのおのの施設で診療に用いられているものとは異なる可能性がある。標準化された方法を使用することにより、検査を実施した施設にかかわらず、個々の研究内および研究間での比較が可能になる。

### 具体的注意

胸部X線は、深吸気でPA方向の撮影(posteroanterior projection)を行うべきであるが、それだけでなくフィルムと管球の距離をどの検査でも一定に保つべきである。しかし、進行疾患に対する試験の患者では、これらの条件を達成できるほど状態がよくない可能性があり、その際は測定値とともに状況を報告する。

胸壁に接する病変は、胸部X線による測定には不適である。患者体位のわずかな変化によって病変の投影面に大きなずれが生じ、アーチファクトによる見かけ上の変化が生じる可能性があるからである。これら

の病変に対しては、CT または MRI を施行しなければならない。同様に、縦隔に接したり浸潤している病変は CT または MRI にて評価しなければならない。

胸部、腹部、骨盤内の CT スキャンでは、当該の解剖学的領域全体を連続的にスキャンする必要がある。おおよその指針としては、最小の病変の大きさはスライス厚の 2 倍以上でなければならない。病変がこれより小さいと、実質的な「部分体積」効果 (partial volume effect) の影響を受けやすくなる [ 断面と最長径のあいだの距離が離れていると大きさが過小評価され、そのような病変は実際は同じ大きさのままなのに次の検査で奏効あるいは増悪しているようにみえることがある ( 図 1 ) ]。ベースライン検査でのスライス厚に応じて病変の大きさの下限を決めることにより、次の検査で小さくなった病変は真に大きさが減少していることが確実となる。個々の標的病変の最長径は水平断面 ( 体軸と直行する面 ) のみで選択する。

スライス厚および病変の最小径に関しては、CT スキャナーの型式が重要になる。ヘリカル CT スキャナーにて 5 mm 間隔で連続再構成した画像を得た場合、ベースラインにおける病変の最小径は 10 mm となる。従来型 CT スキャナーでは、スライス厚 10 mm の連続スキャンでの病変最小径は 20 mm である。

ヘリカル CT と従来型 CT の本質的な違いとしては、後者がスキャンされた特定のスライス厚のみの情報を収集し、グレースケール画像としてその二次元画像を描出すること、またそのスライス厚の画像化が可能になるには次のスライス厚まで走査する必要があることなどがある。ヘリカル CT は典型的には、約 20 ~ 30 秒間の 1 回の息止め下に胸部あるいは上腹部全体の全体積のデータを収集して画像化する。画像を描出する際には、データが正確に画像化されるように、ヘリカル CT 側で適切な再構成アルゴリズムを選ぶ。上述したように、ヘリカル CT では 5 mm 厚の再構成が可能であり、それによって 10 mm を病変の最小径とすることが可能となる。

ヘリカル CT は現在、米国、ヨーロッパ、日本においてがんの治療に携わる多くの施設で標準となっており、したがってヘリカル CT に関連する上記コメントは妥当といえよう。臨床試験を行っている施設には、従来型 CT を用いているものもあるが、従来型 CT のスキャナーの台数はヘリカル CT への転換に従って減

少するであろう。

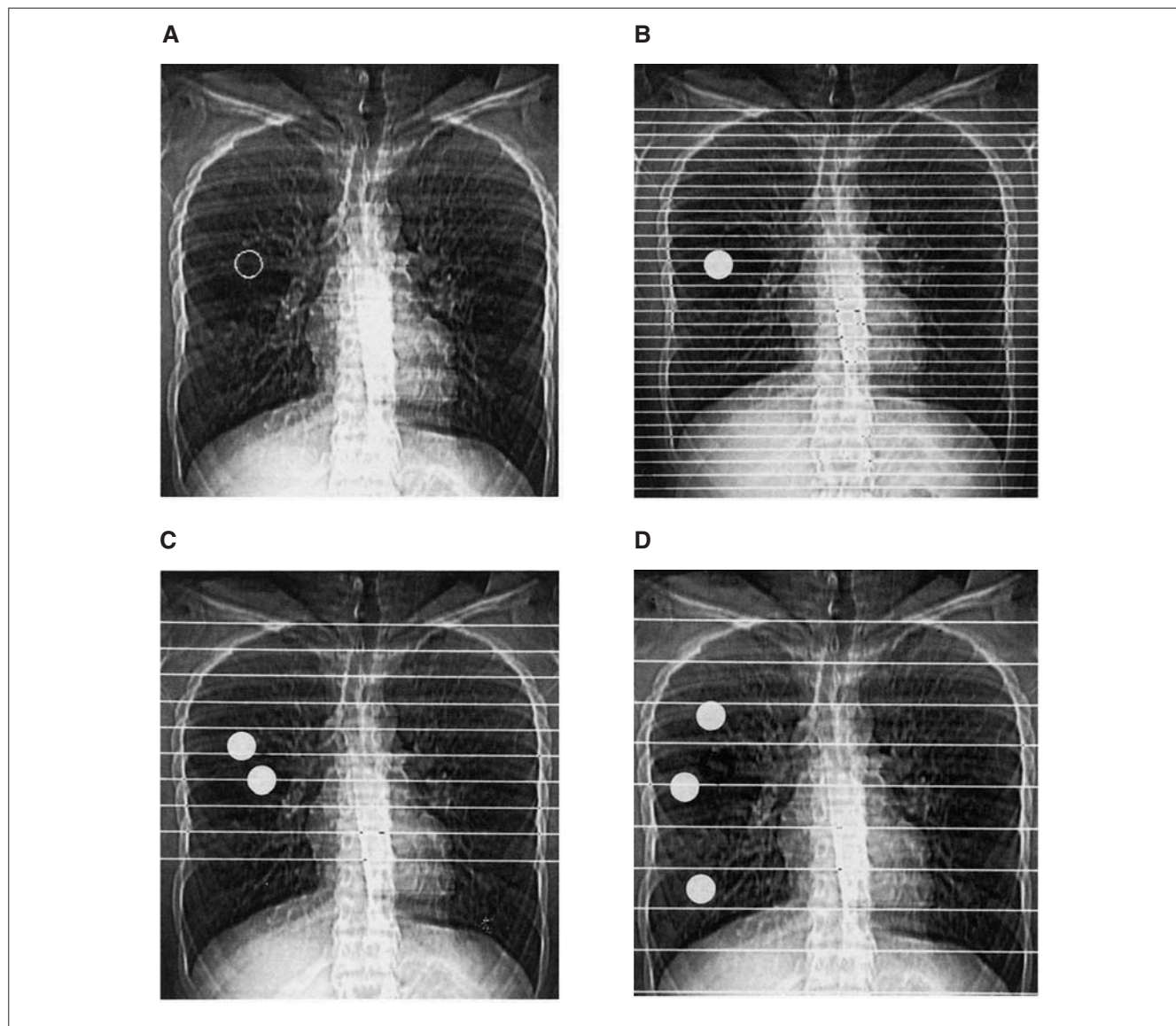
CT スキャンのスライス厚が異なる他の身体部位 ( 典型的には 5 mm 厚が標準である頸部など )、あるいはスライス厚がそれぞれ異なる可能性のある小児科領域では、測定可能な病変の最小の大きさが異なるが、その場合も、スライス厚の 2 倍であるべきである。スライス厚および病変の最小径を試験プロトコールで規定すべきである。

腹部および骨盤内の撮影を行う患者では、他の軟部組織腫瘍に対して腸管を強調するために経口造影剤を投与すべきである。この処置は、ほとんどあらゆる施設でルーチンに実施されている。

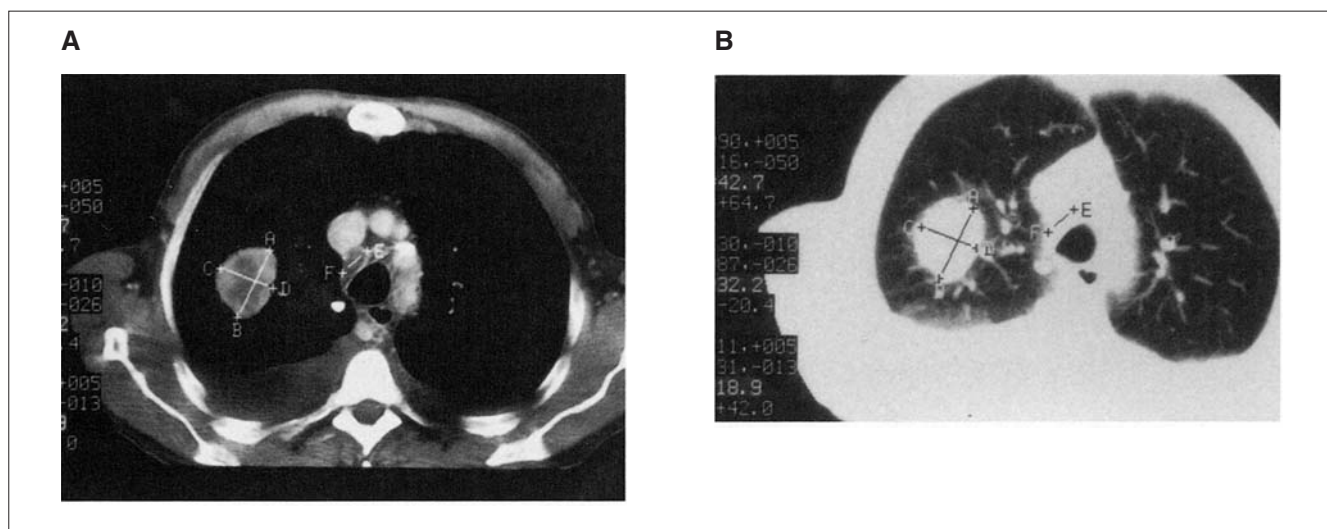
アレルギーなどの医学的理由による禁忌がなければ、静注造影剤も使用すべきである。これは隣接するリンパ節腫瘍に対し血管構造を強調したり、肝などの実質臓器の転移巣の強調に役立つ。日常診療においては、静注造影剤を使用しても新たな情報はほとんど得られない可能性があるが、測定可能な疾患を対象として客観的な奏効率をエンドポイントとする臨床試験では、もしも静注造影剤を用いなければ、用いた場合に測定可能な病変のうち相当数が測定不能となるであろう。特定の疾患部位では、静注造影剤が不要に思われる場合がある ( たとえば、疾患が明らかに肺末梢部に限局している患者 )。しかし、臨床試験では、病変が実際に消失し、他のスキャン部位 ( たとえば、肝の小転移巣 ) に新たな病変が出現した証拠がないことを確実にすることは重要であり、静注造影剤を使用すれば、こうした病変がより容易に描出できる可能性があるため、このような状況においても静注造影剤の使用を考慮すべきである。

静注造影剤の投与法はさまざまである。造影剤の投与法および投与量の厳密な規則を決めようとするよりも、転移巣をもっとも効果的に描出するのに適切な量の造影剤を用いることと、おのおのの患者に施行される一連の検査に際して一貫した方法を用いることを指示することが適切である。

評価には個々の検査のすべての画像を含めるべきであり、明瞭な病変の画像のみを「選択」すべきではない。その理由は、審査 ( review ) が行われる場合に、審査者 ( reviewer ) がほかには異常がないと判断しかねないからである。すべてのウィンドウ設定を含めるべきであり、肺野条件画像と縦隔条件画像 ( 軟部組織条



**図 1** A) 右中肺野に模擬的な 20 mm 病変を有する胸部の CT「スキャノグラム」。  
 B) 10 mm 厚の連続スライスによる胸部 CT「スキャノグラム」。スライス厚内の個々の体積がスキャンされ、平均減衰係数 [ 多数の小立方体(ボクセル)の密度 ] が、グレースケール上の断面画像として二次元(ピクセル)にて空間的に表示される。図中の各線は、当該スライス厚を通過する構造物の平均密度の空間的な表現であり、線は当該レベルでそれを通る薄い「断面」を表すのではない点に留意することが重要である。したがって、平均値が実質的には低下しないくらいに十分な量の病変があるため、20 mm 以上の病変は最低でも 1 枚の画像上にその真の直径が描出されることになる。  
 C) 15 mm 間隔で施行された CT スキャノグラム。スライス厚内の腫瘍量に依存して、平均密度が実質的に過小評価される可能性があり、2 つの病変のうち上の病変ではその傾向が強いが、下の病変では真の腫瘍径に近くなる。この図はプロセスの過剰な単純化ではあるが、CT 再構成の物理学を詳論せずとも要点が明らかになる。  
 D) 24 mm 間隔、10 mm 厚で施行された CT スキャノグラム。病変は直径が描出されたり、部分的に描出されたり、あるいはまったく描出されないこともある。これはきわめて小さな病変を画像化する場合や、ある検査から次の検査のあいだに真の直径が変化したか否の検討を行う場合に相当する。



**図 2** **A)**「軟部組織」ウィンドウ(縦隔条件)による気管分岐部レベルの胸部 CT スキャン。2つの病変をキャリパスで測定した。肺実質内病変の最長径および垂直方向の最長距離を二次元的に測定した。一次元測定では最長径のみの測定が求められる。前縦隔リンパ節は一次元で測定された。  
**B)**「肺」ウィンドウ(肺野条件)で画像化された A と同一の画像で、キャリパスは軟部組織の測定値のまま残してある。肺病変の大きさが異なって描出されている。前縦隔リンパ節はこのウィンドウでは測定できない。どの病変を測定するにせよ、一連の検査では同一のウィンドウを使用すべきである。軟部組織ウィンドウ(縦隔条件)が若干好ましく、傍気管、前縦隔、気管分岐部下の病変などは肺実質内病変としてこの設定でフォローアップするのがよいと思われる。

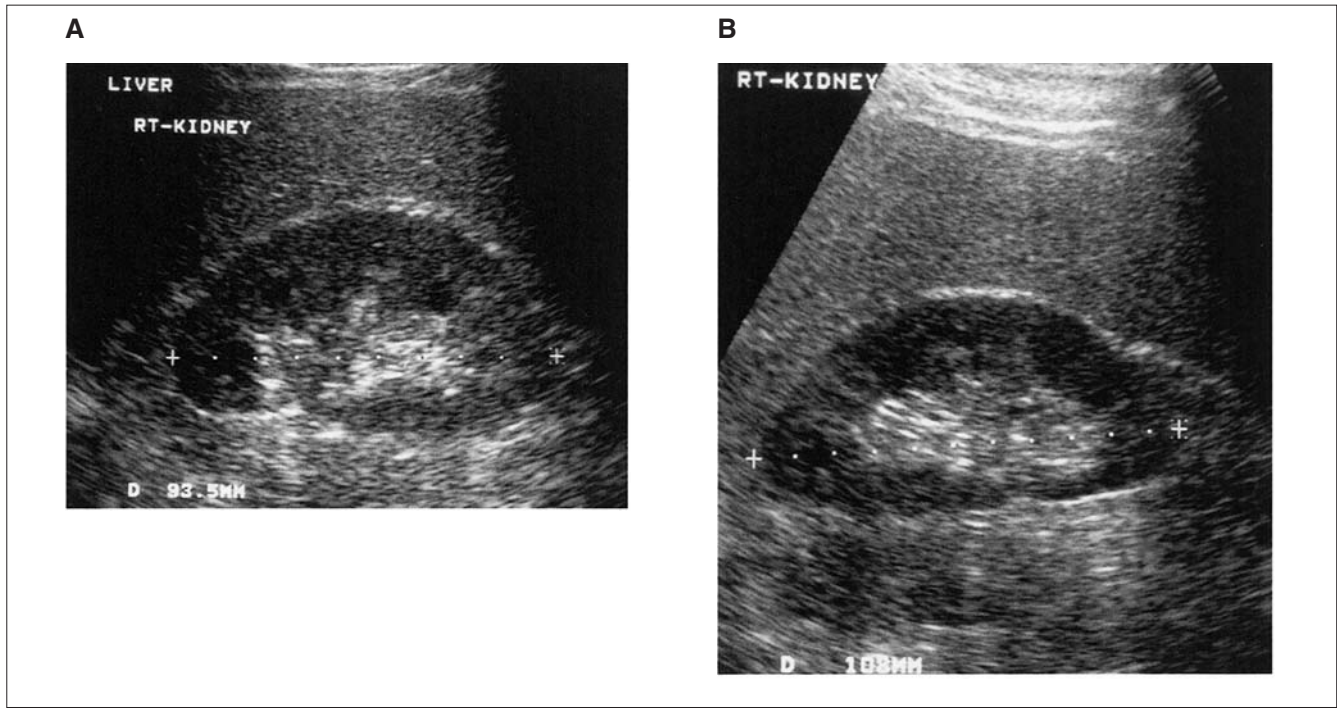
件)を評価すべき胸部についてはとくにそうである。

病変は、個々の検査において同一のウィンドウ設定の画像に基づいて測定されるべきである。1つの病変をある検査では肺野条件画像で測定し、次の検査では縦隔条件画像で測定することは許容されない(図2)。肺では、リンパ節および肺実質の病変の綿密な評価が行われ、試験期間中を通じて反復される検査において同一のウィンドウ設定を用いて標的病変が適切に測定されていれば、肺実質内の病変に対し肺野条件か縦隔条件のいずれが用いられているかは現実的には問題とならない。

MRIの使用には複雑な問題がある。MRIはすべての臓器について種々の解剖学的断面の画像を描出することができる。それゆえに重要なのは、MRIを用いる際、一連の検査で同一の画像シーケンスを用いた同一の解剖学的断面において病変を測定する必要があることである。MRIスキャナーは多彩な画像を作成する。画像の作成に関連する因子には、磁気強度(高磁場のものではスキャン時間がより短く、典型的には2~5分)、コイルのデザイン、患者の協力などがある。可能であれば、同一のスキャナーを用いるべきである。たとえば、1.5テスラのスキャナーによる画像

は0.5テスラのスキャナーの画像とは異なる。異なるスキャナーの画像でも比較は可能であるが、そのような比較は理想的とはいえない。さらに、進行した悪性疾患を有する多くの患者は、疼痛があるため走査シーケンス中(2~5分を要する)に静止状態を維持することができない。スキャン時間中のあらゆる動きが体動アーチファクトと画質の低下をもたらすため、検査が無駄になる可能性が高いと考えられる。これらの理由により、現時点ではCTが選択されるべき画像モダリティである。

超音波検査は主観的となることが避けられないため、臨床試験において、表在性ではない腫瘍の縮小あるいは増大の測定には使用すべきではない。後日行われる独立した第三者による審査のために検査全体を再現することは不可能であり、それが真実かどうかは別として、使用されるハードコピーフィルムはイベントの正確な反映であり真実を再現する、という仮定を置かざるを得なくなる(図3)。さらに、たとえば、ただ1つの測定可能病変が腹部の傍大動脈領域にあって、腸内ガスが病変の上を覆っている場合は、超音波ビームはガスを透過しないため病変は描出されない。したがって、当該患者の疾患のステージング(あるい



**図 3** **A)** 正常構造の右腎の超音波スキャン。キャリパスで 93 mm と測定された。  
**B)** 数分後に 108 mm と測定された同一の腎の超音波スキャン。大きさが 16% 増大している。この差はパネル A の腎の歪みのためである。解剖学的な目印がないため、一連の検査において同一平面上で正確に測定することが困難になる。フィルムのハードコピー上の測定値がイベントの真に正確な反映であることを願う以外にない。

は治療効果の評価のための再ステージング)は正確ではなくなる。

病変の測定には、試験を通じて同一の画像モダリティを用いるべきである。異なる画像技術は感度に差があるため、同じ病変を同じときに測定したとしても、異なるモダリティで測定すれば、測定法によって計測値が異なる可能性がある。したがって、試験中に別のモダリティに変更したり、異なるモダリティの測定値を採用することは許容されない。全体を通じて同一の技術である必要がある。

患者が臨床試験から無用に除外されないようにするためには、必要以上の制約を加えることなく画像モダリティを標準化しようという姿勢が望まれる。

### 付録 II . 直径, 積(面積), 体積の変化とその関係

**付録 II 表 2** 直径, 積(面積), 体積の変化とその関係\*

	直径 ( $2r$ )	積 [ $(2r)^2$ ]	体積 ( $4/3 r^3$ )
奏効 ( response )	減少	減少	減少
	30%	50%	65%
	50%	75%	87%
増悪 ( disease progression )	増加	増加	増加
	12%	25%	40%
	20%	44%	73%
	25%	56%	95%
	30%	69%	120%

\* 網かけ部分は、奏効および増悪の定義を満たす腫瘍の大きさの変化に対する RECIST(直径)と WHO(積)の判定規準を表す。

### 付録 III . RECIST ワーキンググループ および謝辞

#### RECIST ワーキンググループ

P. Therasse (Chair), J. Verweij, M. Van Glabbeke, A. T. van Oosterom, European Organization for Research and Treatment of Cancer (Brussels, Belgium); S. G. Arbuck, R. S. Kaplan, M. C. Christian, National Cancer Institute, United States (Bethesda, MD); E. Eisenhauer, National Cancer Institute of Canada Clinical Trials Group (Kingston); S. Gwyther, East Surrey Hospital (Redhill, U.K.); and J. Wanders, New Drug Development Office Oncology (Amsterdam, The Netherlands).

#### レトロスペクティブ解析

L. A. Rubinstein, National Cancer Institute, United States; B. K. James, A. Muldal, W. Walsh, National Cancer Institute of Canada Clinical Trials Group; S. Green, Southwest Oncology Group (Seattle, WA); M. Terenziani, National Cancer Institute (Milan, Italy); D. Vena, Emmes Corporation (Rockville, MD); R. Canetta, J. Burroughs, Bristol-Myers Squibb (Wallingford, CT); A. Riva, M. Murawsky, Rhone-Poulenc Rorer Pharmaceuticals Inc. (Paris, France).

### 付録 IV . RECIST の最終稿作成ワーク ショップ (1998 年 10 月) 参加者 および謝辞

#### 参加者

S. C. S. Kao, Children's Cancer Study Group (Iowa City, IA); D. Grinblatt, Cancer and Leukemia Group B (CALGB) (Chicago, IL); B. Giantonio, Eastern Cooperative Oncology Group (ECOG) (Philadelphia, PA); F.B. Stehman, Gynecologic Oncology Group (GOG) (Indianapolis, IN); A. Trotti, Radiation Therapy Oncology Group (Tampa, FL); C. A. Coltman, Southwest Oncology Group (SWOG) (San Antonio, TX); R. E. Smith, National Surgical Adjuvant Breast and Bowel Project (Pittsburgh, PA); J. Zalcberg, Peter MacCallum Cancer Institute (Melbourne), Australia; N. Saijo, National Cancer Center Hospital (Tokyo, Japan); Y. Fujiwara, National Institute of Health Sciences (Tokyo); G. Schwartzmann, Hospital de Clinicas de Porto Alegre (Brazil); A. Klein, Health Canada, Bureau of Pharmaceutical Assessment (Ottawa, ON); B. Weirnerman, National Cancer Institute of Canada Clinical Trials Group (Kingston, ON); D. Warr, Ontario Cancer Institute/Princess Margaret Hospital (Toronto); P. Liati, South Europe New Drugs Organization (Milan, Italy); S. Einstein, Bio-Imaging Technologies (West Trenton, NJ); S. Négrier, L. Ollivier, Fédération Nationale

des Centres de Lutte contre le Cancer (Paris, France); M. Marty, International Cancer Cooperative Group/French Drug Agency (Paris); H. Anderson, A. R. Hanauske, European Organization for Research and Treatment of Cancer (EORTC)(Brussels, Belgium); M.R. Mirza, Odense University Hospital (Denmark); J. Erbsoll, The European Agency for the Evaluation of Medicinal Products (Bronshoj, Denmark); C. Pagonis, Cancer Research Campaign (London, U.K.); S. Hatty, Eli Lilly and Co., (Surrey, U.K.); A. Riva, Rhone-Poulenc Rorer Pharmaceuticals Inc. (Paris); C. Royce, GlaxoWellcome (Middlesex, U.K.); G. Burke, Novartis Pharma AG (Basel, Switzerland); I. Horak, Janssen Research Foundation (Beerse, Belgium); G. Hochtin-Boes, Zeneca (Macclesfield Cheshire, U.K.); C. Weil, Bristol-Myers Squibb (Waterloo, Belgium); M.G. Zurlo, Pharmacia & Upjohn (Milan); S. Z. Fields, SmithKline Beecham Pharmaceuticals (Collegeville, PA); B. Osterwalder, Hoffmann-La Roche Inc. (Basel); Y. Shimamura, Taiho Pharmaceutical Co., Ltd. (Tokyo); and M. Okabe, Kyowa-Hakko-Kogyo Co., Ltd. (Tokyo).

#### 追加コメント寄稿者

A. Hamilton, R. De Wit, E. Van Cutsem, J. Wils, J.-L. Lefèbvre, I. Vergote, M. S. Aapro, J.-F. Bosset, M. Hernandez-Bronchud, D. Lacombe, H. J. Schmoll, E. Van Limbergen, P. Fumoleau, A. Bowman, U. Brunsch, EORTC (Brussels); B. Escudier, P. Thiesse, N. Tournemaine, P. Troufleau, C. Lasset, F. Gomez, Fédération Nationale des Centres de Lutte contre le Cancer (Paris); G. Rustin, Mount Vernon Hospital (Northwood Middlesex, U.K.); S. B. Kaye, Western Infirmary (Glasgow, U.K.); A. Goldhirsch, F. Nolè, G. Zampino, F. De Braud, M. Colleoni, E. Munzone, T. De Pas, International Breast Cancer Study Group and Istituto Europeo di Oncologia (Milan); M. Castiglione, J. F. Delaloye, A. Roth, C. Sessa, D. Hess, B. Thürlimann, C. Böhme, T. Cerny, U. Hess, Schweizer Arbeitsgemeinschaft für Klinische Krebs-forschung (Bern, Switzerland); H. J. Stewart, Scottish Cancer Therapy Network (Edinburgh, U.K.); A. Howell, J. F. R. Robertson, United Kingdom Coordinating Committee on Cancer Research (Nottingham); K. Noever, Bio-Imaging Technologies (Monheim, Germany); M. Kurihara, Toyosu Hospital, SHOWA University (Tokyo); L. Seymour, J. Pater, J. Rusthoven, F. Shepherd, J. Maroun, G. Cairncross, D. Stewart, K. Pritchard, National Cancer Institute of Canada Clinical Trials Group (Kingston); T. Uscinowicz, Health Canada, Bureau of Pharmaceutical Assessment (Ottawa); I. Tannock, Princess Margaret Hospital (Toronto); M. Azab, QLT Phototherapeutics (Vancouver, Canada); V. H. C. Bramwell, Canadian Sarcoma Group (London); P. O'Dwyer, ECOG (Philadelphia); A. Martin, S. Ellenberg, U.S. Food and Drug Administration (Rockville, MD); C. Chow, D. Sullivan, A. Murgo, A. Dwyer, J. Tatum, National Cancer Institute (Bethesda, MD); R. Schilsky, CALGB (Chicago, IL); J. Crowley, S. Green, SWOG (Seattle, WA); R. Park,

GOG (Philadelphia, PA); V. Land, B. D. Fletcher, Pediatric Oncology Group (Chicago, IL); B. Hillman, University of Virginia (Charlottesville); F. Muggia, New York University Medical Center (New York); C. Erlichman, Mayo Clinic (Rochester, MN); L. H. Schwartz, Memorial Sloan-Kettering Cancer Center (New York, NY); S. P. Balcerzak, Ohio State University Health Sciences Center (Columbus); G. Fleming, CALGB (Chicago); G. Sorensen, Harvard University (Cambridge, MA); H. Levy, Thomas Jefferson University (Philadelphia); N. Patz, Duke University (Durham, NC); C. Visseren-Grul, Eli Lilly Nederland BV (Nieuwegein, The Netherlands)/J. Walling, Lilly Research Laboratories (Indianapolis); P. Hellems, Janssen Research Foundation (Beerse, Belgium); L. Finke, Merck (Darmstadt, Germany); A. Man, N. Barbet, Novartis Pharma AG (Basel); G. Massimini, Pharmacia & Upjohn (Milan); J. Jimeno, Pharma Mar (Madrid, Spain); I. Hudson, SmithKline Beecham Pharmaceuticals (Essex, U.K.); and J. Krebs, R. A. Beckman, S. Lane, D. Fitts, SmithKline Beecham Pharmaceuticals (Collegeville).

## 付録 V . WHO/SWOG 規準および新 RECIST 規準で得られた奏効率/増悪率のレトロスペクティブな比較

腫瘍病変の一次元測定がこれまでの二次元的アプローチに代わりうるという仮説を評価するために、多くのレトロスペクティブな解析が実施された。以下に、これらの解析の結果を示す。

### 1 . WHO 法 (もしくは WHO 変法) あるいは RECIST 法を用いた奏効率と増悪率の比較

#### 1.1. 評価の対象となった試験

試験データに腫瘍測定の一連の(反復)記録が含ま

れること以外には、特別な選択規準は設けなかった。単一あるいは複数の試験に基づいて自前のデータの評価を行った研究グループ [カナダ NCI 臨床試験グループ/キングストン/オンタリオ州, 米国 NCI/ベセスダ/メリーランド州, Rhone-Poulenc Rorer 製薬/パリ/フランス] と、評価に用いるデータを米国 NCI に提供した研究グループ (SWOG および Bristol-Myers Squibb 社/ワリングフォード/コネチカット州) があつた。

#### 1.2. 評価の対象となった効果判定規準

治療効果のすべての指標を評価するためにすべてのデータベースが用いられたわけではない。データ処理開始時には、ほとんどの関心は WHO 規準および新 RECIST 規準による奏効 (CR + PR) 率の比較に向けられていた。これらのデータにより、新規規準を使用しても奏効率に影響を及ぼさないことが示唆されると、さらにいくつかのデータベースを用いて、CR + PR だけでなく SD および PD 率 (付録 V, 表 4 を参照), さらに無増悪期間 (time to disease progression) (付録 V, 表 5 を参照) に及ぼす新規規準使用の影響が解析された。

#### 1.3. 比較の方法

個々の試験の各患者について、ベースラインにおける合計値が計算された (WHO では垂直方向の 2 つの最長径の積の和, RECIST では最長径の和)。個々の評価ののうち、新たに腫瘍の測定が可能な場合は再計算が行われた。付録 V の表 3 に示す測定規準に基づく「最良」効果 (“best” response) として、各患者の CR, PR, SD, PD が決定された。WHO および RECIST とともに、CR, PR の確定を検討するために最低でも 4 週間

付録 V 表 3 WHO あるいは RECIST 規準による最良効果の定義\*

最良効果	WHO (積の和の変化)	RECIST (最長径の和の変化)
CR	消失 (4 週間後に確定 <sup>†</sup> )	消失 (4 週間後に確定 <sup>†</sup> )
PR	50% の減少 (4 週間後に確定 <sup>†</sup> )	30% の減少 (4 週間後に確定 <sup>†</sup> )
SD	PR の規準も PD の規準も満たさない	PR の規準も PD の規準も満たさない
PD	25% の増加 (病変が増大する前に CR, PR, SD と判定されない)	20% の増加 (病変が増大する前に CR, PR, SD と判定されない)

\* WHO = World Health Organization (世界保健機関), RECIST = Response Evaluation Criteria in Solid Tumors (固形がんの効果判定規準), CR = complete response (完全奏効), PR = partial response (部分奏効), SD = stable disease (安定), PD = progressive disease (進行)

<sup>†</sup> Bristol-Myers Squibb 社 (ワリングフォード/コネチカット州) のデータセットのうち、一次元測定 (RECIST 規準) による最良効果と二次元測定 (WHO 規準) による最良効果との比較には、未確定の CR および PR のみが用いられた。このデータセットでは、効果の確定を示すコードが、技術的な理由から比較に用いることができなかった。



付録 V 表 4 14 試験に登録された同一患者における RECIST(一次元)判定規準と WHO(二次元)判定規準の比較\*

腫瘍部位/がん種	判定規準	評価症例数	最良効果				奏効率	PD 率
			CR	PR	SD	PD		
乳がん <sup>†</sup>	WHO	48	4	22			54%	
	RECIST	48	4	22			54%	
乳がん <sup>‡</sup>	WHO	172	4	36			23%	
	RECIST	172	4	40			26%	
脳腫瘍 <sup>†</sup>	WHO	31	12	10			71%	
	RECIST	31	12	10			71%	
メラノーマ <sup>†</sup>	WHO	190	9	37			24%	
	RECIST	190	9	34			23%	
乳がん <sup>§</sup>	WHO	531	50	102			29%	
	RECIST	531	50	108			30%	
結腸がん <sup>§</sup>	WHO	1,096	12	137			14%	
	RECIST	1,096	12	133			13%	
肺がん <sup>§</sup>	WHO	1,197	60	317			32%	
	RECIST	1,197	60	318			32%	
卵巣がん <sup>§</sup>	WHO	554	24	108			24%	
	RECIST	554	24	105			23%	
肺がん <sup>†</sup>	WHO	24	0	4	16	4	17%	17%
	RECIST	24	0	4	19	1	17%	4%
結腸がん <sup>†</sup>	WHO	31	1	6	15	9	23%	29%
	RECIST	31	1	5	16	9	21%	29%
肉腫 <sup>†</sup>	WHO	28	1	4	13	10	18%	36%
	RECIST	28	1	5	17	5	21%	18%
卵巣がん <sup>†</sup>	WHO	45	0	7	19	19	16%	42%
	RECIST	45	0	6	21	18	13%	40%
乳がん	WHO	306	18	114	117	57	43%	19%
	RECIST	306	18	108	124	56	41%	18%
乳がん	WHO	360	10	73	135	142	23%	39%
	RECIST	361	10	70	139	142	22%	39%
合計 (腫瘍縮小効果の評価が行われた全例)	WHO	4,613	205	977			25.6%	
	RECIST	4,614	205	968			25.4%	
合計 (CR + PR のほか PD の評価が行われた全例)	WHO	794			315	241		30.3%
	RECIST	795			336	231		29%

\* WHO = World Health Organization (世界保健機関<sup>§</sup>), RECIST = Response Evaluation Criteria in Solid Tumors (固形がんの効果判定規準), CR = complete response (完全奏効), PR = partial response (部分奏効), SD = stable disease (安定), PD = progressive disease (進行)

<sup>†</sup> カナダ NCI 臨床試験グループによる第 II/III 相試験のデータ

<sup>‡</sup> 米国 NCI による第 III 相試験のデータ

<sup>§</sup> Bristol-Myers Squibb 社 (ワリントン/コネチカット州) による第 II/III 相試験のデータ

Rhone-Poulenc Rorer 製薬 (パリ/フランス) による第 III 相試験 (注: 本データベース中には病変の測定が一次元でのみなされた症例が 1 例あり, WHO 判定規準による評価は不能であった。)

付録 V 表 5 個々の評価法による増悪症例の割合\*

	症例数	%
全増悪症例数	234	100
新病変の出現による増悪 <sup>†</sup>	118	50
既存の測定可能病変の増大による増悪	116	50
WHO および RECIST 判定規準による増悪判定日が同じ	215	91.9
WHO および RECIST 判定規準による増悪判定日が異なる	19	8.1
WHO 判定規準でより早期に PD	17	7.3
RECIST 判定規準でより早期に PD	2	0.9

\* PD = progressive disease (進行), WHO = World Health Organization (世界保健機関), RECIST = Response Evaluation Criteria in Solid Tumors (固形がんの効果判定規準)

<sup>†</sup> 測定不能病変の明らかな増大による PD 症例が数例含まれる。

付録 V 表 6 増悪判定日の不一致がある場合の差の程度\*

	症例数	%(234例に対する割合, 表5を参照)
増悪判定日が異なる増悪症例数	19	8.1
8~9週の差	3	1.3
12週の差	1	0.4
24~31週の差 <sup>†</sup>	2	0.9
WHO, RECIST どちらかによる増悪判定日の打ち切りにより差が不明 <sup>‡</sup>	13	5.6

\* WHO = World Health Organization (世界保健機関), RECIST = Response Evaluation Criteria in Solid Tumors (固形がんの効果判定規準)

<sup>†</sup> 1例では、増悪が一次的であったことが主な理由で、RECIST 判定規準(一次元)による増悪判定が WHO 判定規準よりも24週早かった。横断面が25%増大した結腸がんの1例は、いったん寛解したのち再発したため、WHO 判定規準による増悪判定が RECIST 判定規準よりも31週早かった。

<sup>‡</sup> 本表に示すように、19例中13例において RECIST と WHO 判定規準で比較した場合に増悪日の差が不明であった。これらの症例では、RECIST での増悪の判定が SWOG 判定規準<sup>5)</sup>で増悪(腫瘍断面が50%あるいは10 cm<sup>2</sup>増大)とされる条件を満たさなかった。とくに、このうち6例が WHO(二次元的に25%の増大)および SWOG(二次元的に50%の増大)の規準で判定された増悪日が同じであった。一次的な20%の増大(RECIST)は二次元的には約44%の増大に相当するため、確実ではないが、SWOG および WHO 判定規準で増悪とされてからまもなく RECIST の一次的な判定規準による増悪が起こる傾向がある。3例では、WHO と SWOG 判定規準で二次元的に50%の増大と判定される日に10~12週の差がみられた。やはり同じく証明はできないが、その後まもなく RECIST 判定規準が満たされる傾向がみられる。WHO と RECIST 判定規準による増悪の時期の差が不明であった13例のうち残りの4例は、腫瘍の表面積の増大が10 cm<sup>2</sup>以上であったため SWOG 判定規準<sup>5)</sup>に従って増悪(PD)に分類された。これらの症例では差の程度がまったく不明である。

の間隔を置くことが求められた。すなわち、個々の患者は2つの規準に従ってそれぞれ最良効果が判定された。奏効率(overall response rate)および増悪率(disease progression rate)は、個々の試験の対象集団あるいは用いられたデータセットをもとに算出された。

(注: WHO 規準の PD に対しては、多くの研究グループでの慣行に従い単一病変の増大ではなく、積の和の増加が用いられた。)

#### 1.4. 結果

付録 V 表 4 に示す。

## 2. 無増悪期間の評価

無増悪期間(time to disease progression)は、SWOG 提供のデータセットをもとに WHO 規準を RECIST と比較することで評価した。増悪に関する SWOG 規準<sup>5)</sup>は、積の和の50%の増加または新病変の出現、または積の和の絶対値の10 cm<sup>2</sup>増大であるため、このデータセットを用いて、積の和の25%の増加と最長径の和の20%の増加(積の和の約44%の増加に等しい)からそれぞれ得られる無増悪期間の差を評価することができた(表5, 表6参照)。

### 2.1. 評価の対象となったデータセット

データセットには、SWOG<sup>5)</sup>の定義による PD 患者

234例が含まれた。全例ともベースラインにおいて測定可能病変を有し、その後同じ方法を用いて PD と判定された。本データセットに含まれたがん種は、メラノーマ、大腸がん、肺がん、乳がんであった。

## References

- (1) Zubrod CG, Schneiderman SM, Frei E III, Brindley C, Gold GL, Schnider B, *et al.* Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and thio phosphamide. *J Chronic Dis* 1960;**11**:7-33.
- (2) Gehan E, Schneidemann M. Historical and methodological developments in clinical trials at the National Cancer Institute. *Stat Med* 1990;**9**:871-80.
- (3) WHO handbook for reporting results of cancer treatment. Geneva (Switzerland): World Health Organization Offset Publication No. 48; 1979.
- (4) Miller AB, Hogestraeten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer* 1981;**47**:207-14.
- (5) Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. *Invest New Drugs* 1992;**10**:239-53.
- (6) James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Mudal A, *et al.* Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst* 1999;**91**:523-8.

## Note

原稿受領日: 1999年6月24日, 改訂原稿受領日: 1999年11月30日, 掲載承認日: 1999年12月3日