

では、OSをprimary endpointとすべき」と結論している⁵⁾。このrecommendationに従えば、OSに対するsurrogacyが示されておらず、PFSがtrueともいえないがん種では、当面OSをprimary endpointとして使い続けるしかないという身も蓋もない結論になる。

現実的な解決策はあるか？

しかし、現実問題として、OSが20か月以上になるようながん種では、OSをprimary endpointとしてランダム化比較試験を組むことは、登録数の観点から困難である場合も多い。たとえばOSの3か月の延長が「臨床的に意味のある差」である場合、6か月に対して3か月の上乘せを検証するには212例でよかったが、これが12か月に対する3か月の上乘せだと838例が必要になり、18か月だと2,092例、24か月だと4,194例が必要となる(話を単純化するために片側 $\alpha=0.025$, $\beta=0.2$, 登録3年, 追跡1年に固定)。理論的にOSが望ましいというのは簡単であるが、ではどのようにしてこのような難しい状況のもとで治療開発を進めていくのか、ということを実験臨床の実務家としては考える必要がある。

PPSが延長するに従って増え続けるサンプルサイズに対する対処法としては、次にあげる3通りの方策が考えられる。

1. 大規模なサンプルサイズに対応できる体制を構築する

最初の選択肢は最もストレートな解決策であるが、単一の臨床試験グループでは困難なサンプルサイズの試験を、共同試験という形を取ることで実施可能にするというアイデアである。近年、国際共同試験や国内共同試験の必要性が叫ばれているが、共同試験の重要な動機の一つは、OSをprimary endpointとした時に増え続けるサンプルサイズである。ほかにも共同試験の動機には、個別化治療が進むに従ってマーカー別に細分化される試験の対象に対して迅速に患者登録を進めるといったことや、(試験の場合は)同時治療開発によるコストダウンや、結果を一般化できる地域を広げたいといった要因も存在するが、いずれにしても単にお友達を増やそうというnaïveな感情が共同試験の動機ではないこ

とを理解しておく必要はあるだろう。一方で、国際共同試験では国別にアウトカム差があることが知られておりデメリットも存在する(単に人種間差ということだけではなく、適格性の診断精度の違いや、保険制度の違いによる後治療の差など多くの要因が原因となりうる)。これらのデメリットが存在することから、比較的文化的習慣が似た地域のグループと共同試験を行えるような体制作りが理想的で、それらのグループとお互いの信頼関係を醸成するとともに実施体制の標準化を進めるべきであろう。

2. デザインの工夫によるサンプルサイズの調整

(1) 試験の精度を落とす

試験対象となる集団の予後は変えることができないため、通常は優越性デザインを取る限り、サンプルサイズを減らすために緩められるパラメーターは、 α (有意水準), β (本当は有効な治療を捨ててしまう確率, $1-\beta$ が検出力), Δ (臨床的に意味のある差)の3つしかない。 Δ はこれぐらい差があれば、試験治療の毒性などのデメリットを考慮に入れても試験治療を新しい標準と見なすうるといった臨床的な感覚を反映した数字であるため、この Δ を緩めてしまうと臨床的に意味のある差でもnegativeになってしまう。そのため臨床的な意思決定と試験結果を一致させる観点からは Δ を緩めることはおすすめできない。

β を緩めるということは検出力を落とすということであり、本当は試験治療が良い時に、試験結果としてはnegativeとなってしまう確率が増えてしまうことを意味する。もちろん β も小さい(検出力が高い)ことに越したことはないのだが、他の2つのパラメーターの優先度が高いため、一定の効果が持つ標準治療が存在する状況では先に β が犠牲にされるケースが多い。

α を緩めるということは、本当はよくない治療法をpositiveと結論してしまう危険性を高めることであり、患者に効果のない治療法を還元することを避けるという観点では α を大きくすべきではない。ただし、試験全体の α (有意水準)は両側5%もしくは片側2.5%が世界標準であるが、対象疾患の患者数が限られている場合には片側5%や、かなりの希少がんである場合には片

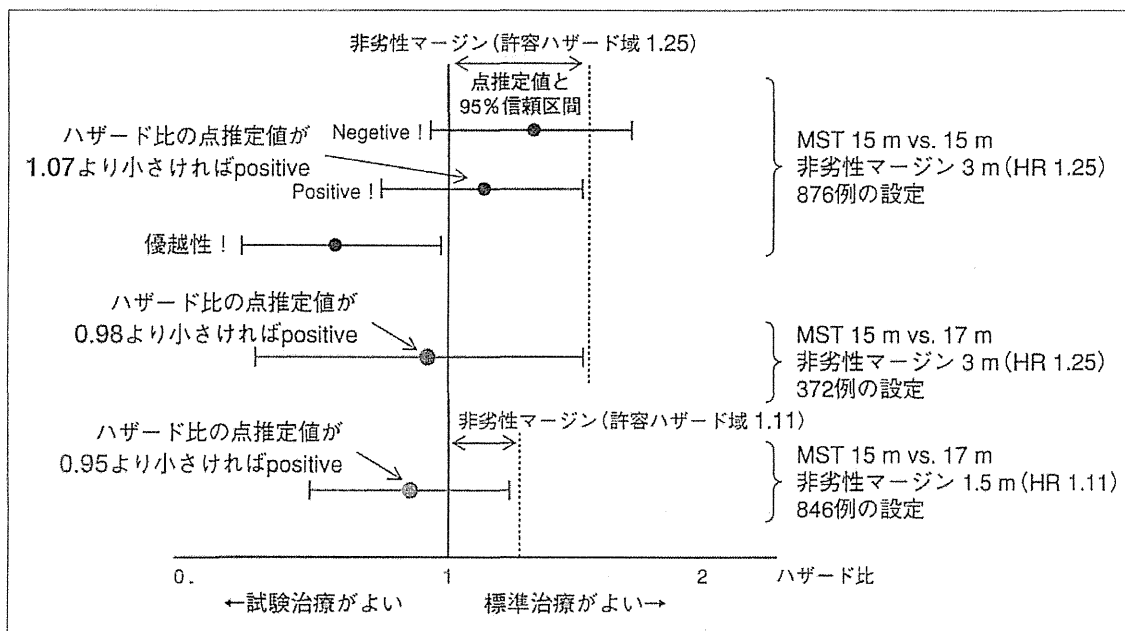


図3 “Hybrid design”と判断境界(boundary)

側10%も許容されるかもしれない。以上のような観点から、一般論として、サンプルサイズが実現不可能なほど多い場合には、まず β 、次に α を緩めることが奨められ、 Δ を緩めることは最後まで避けるべきである。

(2) “Hybrid design”

ある特殊な状況では、NCIのFreidlinらが“hybrid design”と呼んでいる非劣性試験の設定を行うことにより実現可能なサンプルサイズ設定を行える場合がある⁹⁾(なお、世の中にはさまざまな“hybrid design”が存在し、このデザインだけが一意に“hybrid design”と呼ばれているわけではないことには注意が必要である。また、これは特殊なデザインではなく、単に一般的な非劣性試験の標準治療群と試験治療群で見込まれる予後の設定を変えただけにすぎない)。

非劣性試験では特に事情がない限りは標準治療と試験治療で見込まれる予後は同じに設定される。たとえば標準治療群と試験治療群で見込まれる生存期間中央値(median survival time; MST)がどちらも15か月で、非劣性マージンが3か月といった具合である。この時片側有意水準2.5%、検出力80%、登録3年、追跡1年とするとサンプルサイズは876例となる。これに対して“hybrid design”では少し試験治療の予後が上回

るという仮定を置く。たとえば、標準治療群で見込まれるMSTが15か月、試験治療群で見込まれるMSTが17か月、非劣性マージンが3か月といった具合である。先ほどと同じ有意水準、検出力などを用いた場合にこの設定だとサンプルサイズは372例にまで減る。

ただし、サンプルサイズが減った分、得られた結果(ハザード比)の信頼区間の幅は広くなり、positiveになるために越えるべきハードル(判断境界=boundaryという)は高くなる(図3)。たとえばMSTを両群で15か月と見込んだ場合のboundaryは1.07であり、実際に得られたハザード比の点推定値が1.07より小さければP値が2.5%を下回ることになる。つまり試験治療群の生存曲線が少しだけ劣っていてもpositiveになりうる。これに対して標準治療群のMSTを15か月、試験治療群のMSTを17か月(他のパラメータは同じに設定)とした場合のboundaryは0.98となる。つまり試験治療群の生存曲線がわずかではあるが上回っていないとpositiveにはならない。

ちなみに標準治療群のMSTを15か月、試験治療群のMSTを17か月として、非劣性マージンを3か月から1.5か月に狭めた場合、その分サンプルサイズは大きくなり846例と当初の設定とほぼ同じ程度となる。この場合のboundaryは0.95とな

り、試験治療群が少し勝たなければpositiveとされない設定となる。試験治療が標準治療と比較してなんらかのメリットは持つものの、それがさほど大きなメリットではなく、あまり大きな非劣性マージンが許容されない場合などに、このように「少し勝たなければpositiveにならない」という設定にすることはありえる。

この“hybrid design”は設定次第でサンプルサイズを劇的に減らすことができるデザインの方法であるが、適用にあたって2つの注意点がある。まず1点目は、“hybrid design”はあくまで非劣性試験の亜型であり、標準治療に比べて試験治療にOS以外のなんらかのメリットがなければならぬという点である。“Hybrid design”を取るにしても生存曲線がほぼ重なっていた際に“positive”という結論になることは変わらないので、OS以外に毒性が軽い、投与が簡便、といったなんらかのメリットがないことには試験治療を選択することが正当化されない(ただし、先ほどの例のように非劣性マージンがかなり小さい場合には、示すべきメリットも相対的に小さくてよい)。世の中には標準治療と試験治療の毒性プロファイルが異なり、試験治療がtoxic newともless toxicともいえないような、equitoxicともいうべき状況が存在するが、このような場合にも試験治療に少なくともなんらかのメリットが存在する場合には、この“hybrid design”が適用できる可能性がある。

2点目の注意点は、試験治療が少し上回るという仮定を置く以上、過去のデータなどから試験治療が少し勝つ見込みがなければならぬという点である。“Hybrid design”ではサンプルサイズを減らした分、試験がpositiveとなるために越えるべきハードルが高くなるので、本当に勝つ見込みがなければpositiveとなる可能性がその分低くなる。つまり、単なるunderpowerの試験を行っているだけ、ということになりかねない。“Hybrid design”は設定次第でサンプルサイズを劇的に減らすことができるデザインであるが、その適用にあたっては上記の2点がクリアできているか慎重に検討してから適用すべきである。

3. PFSとOSのどちらをもdecision ruleに組み込む

3つ目のオプションとしては、PFSに多少なりともclinical benefitがあることが示される場合に、PFSでサンプルサイズ設定を行い、さらにOSによるdecision ruleを組み込むという方法が考えられる。PFSでサンプルサイズ設定を行うことによって登録数は現実的な範囲内におさめ、試験のdecision ruleとしてはPFSが有意で、かつOSで試験治療が標準治療を一定以上上回った場合に“positive”と結論づけるというデザインである。

喉頭温存生存期間のように誰の目にも明らかなclinical benefitを持つ場合には、OSが重なっていたとしても試験治療を選択するであろうが、ほとんどの場合PFSにclinical benefitがあるといっても、OSのbenefitに比べると相対的に小さいことがほとんどである。そのような状況では、試験治療の真のOSが、標準治療の真のOSを下回っている際に、間違っても試験治療を取ってしまう確率を一定以下に抑えるために、OSが一定以上上回る(例:OSのハザード比の点推定値が0.85を下回る=検定は行わない)というdecision ruleを設ける、というのがこのデザインの基本的なアイデアである。

OSがどれほど上回ればpositiveと結論するかは、喉頭温存生存期間のような明らかなclinical benefitを有するエンドポイントである場合にはこのハードルは低くて済むであろうし、PFSのclinical benefitがそこまで明確でない場合には、このハードルは相対的に高くなる。ただ、少なくともPFSがなんらかのclinical benefitを持つ(trueである)ことが前提であり、そうでなければtrueでもなくsurrogateでもないPFSをdecision ruleに組み込むことが正当化されない。また、OSのdecision ruleについては、 α と β を大幅に緩めた検定を行っていることと同じであるため、この精度が著しく低くないことは確認する必要がある。

Japan Clinical Oncology Group(JCOG)では、まだこのデザインの実例はないが、1.や2.のオプションが適用できない場合に考慮に値するデザインであると考えている。ただし、いまだ広く受け入れられているデザインではないため、設定に際しては生物統計家への相談が必須で

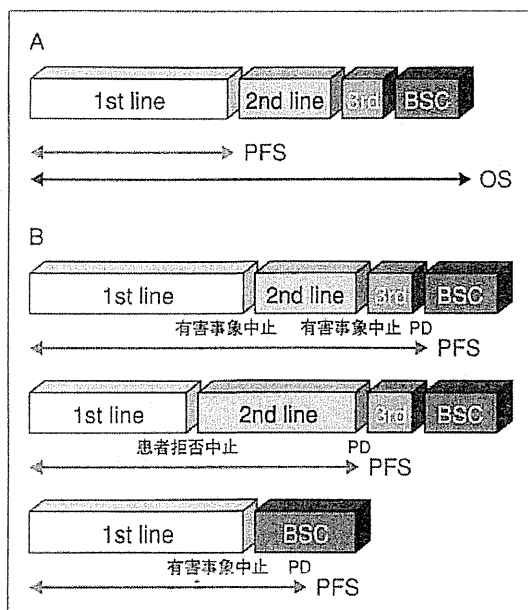


図4 PFSは一次治療の有効性の指標か？

ある。

なお、企業治験などではPFSとOSにたとえば2.5%ずつ有意水準を分割するようなデザインが取られることもあるが、かえってサンプルサイズは大きくなり、PFSがpositiveとなった段階で(OSの結果が出るのは通常その数年後となる)、Food and Drug Administration (FDA) のaccelerated approvalを取りに行くといった企業戦略に基づいたデザインであることから、研究者主導臨床試験で適用できる状況は少ないと考えられる。また、PFSの帰無仮説のハードルをあげる方法も提唱されているが(通常PFSの信頼区間上限が1を下回ればpositiveとなるが、0.9や0.8を下回るべきというdecision rule)⁴⁾、これもPFSがOSに対して一定のsurrogacyを持つことが前提となっており、本稿では割愛する。

HardnessからみたPFS

これまで、truenessの観点から考えた際には、PFSがprimary endpointとして用いることのできる条件は、①PFSがOSのsurrogate endpointとなることが示されている、②PFS自体が患者にとってのclinical benefitを持つ、の2通りしかないということと、状況によって取りうるデザイン上の工夫について述べてきた。ただ、PFSがtrueで

あれ、surrogateであれ、hardであることは必須条件である。PFS評価にはさまざまなバイアスが入りうることは、多くの論文で述べられておりここでは詳述しないが、代表的なものだけでもassessment bias(検査間隔が群間で異なったり、増悪の評価自体が客観的でない場合に生じるバイアス)、attrition bias(主に試験治療群で評価の脱落が多くなるバイアス)⁷⁾、術前療法の試験で生じるlead time bias(手術日のタイミングが群間で異なることで生じるバイアス)⁸⁾といったバイアスがあげられる。注意が必要なのは、ほとんどの場合これらのバイアスは試験治療群に有利に働くという事実であり、PFSをprimary endpointに設定する際には、これらのバイアスを最小化する努力が払われるべきである。

Hardnessの観点に関連して、PFSをめぐる議論であまり意識されていない事実の一つ指摘しておきたい。よくPFSは一次治療の有効性の指標であるという説明がされ、図4-Aのようなスライドが示されることがあるが、これは誤解を招くシェーマである。特に一次治療レジメンの毒性が強い場合などは、プロトコール治療中止理由の半分以上が増悪中止以外の毒性中止や患者拒否中止であることがある。図4-Bに示すように、この場合のPFSのイベント日は一次治療が増悪中止となるまでの期間ではなく、多くのケースで二次治療や三次治療の増悪中止までの期間である(試験によっては、毒性中止や患者拒否中止の時点で「打ち切り」としている場合もある。ただし、打ち切りとすると治療効果の推定にバイアスを生じさせる可能性がある)。純粋な一次治療の治療中止までの期間を評価したいのであれば治療成功期間(time to treatment failure; TTF)を評価すべきであり、TTFとPFSを混同したような議論が行われることがあるため注意が必要である。

また、hardnessという観点では、実は臨床試験グループや試験によってPFSの定義がまちまちであるという事実にも目を向ける必要がある。JCOGにおけるPFSの標準的な定義は画像によるPDと臨床的増悪(symptomatic deterioration)をイベントとし、最終無増悪生存確認日で打ち切りにするというもので、治療中止日や後治療開

始日では打ち切りにしていない。ただ、試験によっては臨床的増悪をイベントにしない定義や（この場合にはしばしばattrition biasが生じる）、後治療開始日で打ち切りにするといった定義、打ち切り日を最終無増悪生存確認日ではなくて最終生存確認日にするといった定義が取られる場合もある。意図的かどうかはわからないが、PFSの結果を提示する際に生存曲線に打ち切りのヒゲを表示せずに発表を行っている場合もあり、ひとくちにPFSといっても定義も質もさまざまであることは念頭に置いた上で結果の解釈を行うべきである。

おわりに

本稿でこれまで行った議論は、“effectiveness”の評価、つまり、患者にとって第一選択として推奨すべき標準治療を確立するランダム化第III相試験を前提としている。新薬に薬効(efficacy)が存在するかどうかを探索する試験では薬効をよりシャープに評価しうるPFSやTTFをprimary endpointとすべき状況はありうるだろうし、研究者主導試験であってもものに第III相試験が控えている第II相試験などのスクリーニング段階ではPFSをもっと積極的に用いるべき状況もありうると思われる。「PFSはprimary endpointとして使えるかどうか？」という議論は十把一絡げにできるものではなく、がん種や試験の相、試験の目的によって変わりうる。ただし、標準治療を確立する第III相試験では本稿で述べたようなtrue-nessやhardnessを十分考察の上で、PFSをprimary endpointとすべきかどうかを決定すべきである。PFSをめぐる議論は往々にして議論のポイントが散漫となり、結論が出ないままに終わることが多いが、本稿で示したフレームワークやデザインの工夫がそれらの議論の一助となれば幸いである。

補遺：なお、本稿の内容は、独立行政法人国立がん研究センターがん研究開発費23-A-16(主任研究者：福田治彦)に基づく研究成果によるものである。

文 献

- 1) Broglio KR, Berry DA. Detecting an overall survival benefit that is derived from progression-free survival. *J Natl Cancer Inst* 2009 ; 101 : 1642.
- 2) Buyse M, Burzykowski T, Carroll K, et al. Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 2007 ; 25 : 5218.
- 3) Shi Q, De Gramont A, Buyse ME, et al. Individual patient data (IPD) analysis of progression-free survival (PFS) versus overall survival (OS) as an endpoint for metastatic colorectal cancer (mCRC) in modern trials : Findings from the 16,700 patients (pts) ARCAD database [abstract]. *J Clin Oncol* 2013 ; 31 Suppl : 3533.
- 4) Fleming TR, Rothmann MD, Lu HL. Issues in using progression-free survival when evaluating oncology products. *J Clin Oncol* 2009 ; 27 : 2874.
- 5) Korn EL, Freidlin B, Abrams JS. Overall survival as the outcome for randomized clinical trials with effective subsequent therapies. *J Clin Oncol* 2011 ; 29 : 2439.
- 6) Freidlin B, Korn EL, George SL, Gray R. Randomized clinical trial design for assessing noninferiority when superiority is expected. *J Clin Oncol* 2007 ; 25 : 5019.
- 7) Freidlin B, Korn EL, Hunsberger S, et al. Proposal for the use of progression-free survival in unblinded randomized trials. *J Clin Oncol* 2007 ; 25 : 2122.
- 8) Nakamura K, Shibata T, Takashima A, et al. Evaluation of three definitions of progression-free survival in preoperative cancer therapy (JCOG0801-A). *Jpn J Clin Oncol* 2012 ; 42 : 896.

第13章

臨床試験の情報提供とコミュニケーション

なぜ臨床試験において情報提供が大切なのか。CRCの役割を研究マネジメント、データマネジメント、患者支援に分けてその理由を考えてみたい。

研究マネジメントは、試験遂行に関係する部門間(病棟・外来・検査部など)や医師、コメディカルとの業務調整やスケジュール管理が挙げられる。スケジュール管理には、個々の患者に関する管理と、試験全体の管理がある。患者ごとのスケジュール管理は、来院や入院の日程の調整、薬物動態など、各種検査の採血時間や食事・試験薬内服のタイミングを説明し、理解してもらうことなどである。

一方、試験全体のスケジュール管理は、試験開始までに必要な準備を整え、あらかじめ、予定された期間内に症例登録が達成できるよう試験全体の進捗管理をすることである。

これらのマネジメントのためにもコミュニケーションスキルは重要である。

データマネジメントは、適切な時期に必要なデータを得て、速やかに提出するための調整となる。そのためには、来院日を守るなど、患者の協力も必要となる。

患者支援には、自己決定の支援と、そのための情報提供や、試験や治療に対する不安等の対応がある。患者からの信頼を得られてこそ必要なデータを得ることにつながる。

情報提供とコミュニケーションには密接な関係がある。

いうまでもないことであるが、CRCとして患者に信頼されるためには、対応なども含めたコミュニケーションスキルが不可欠である。

CRCは、多くの場合、初対面の患者に臨床試験について説明を行い、同意の意思確認を行う。その際には、患者が試験参加について意思決定するための十分でわかりやすい情報提供が必要であり、その過程で患者との間に信頼関係が生まれる。信頼関係が生まれるとさらによいコミュニケーションがとれるようになるのである。

一方で、当然のことではあるが、医師と患者の間には、情報の量や内容の不均衡が存在し、コミュニケーションの障害になることがしばしばある。CRCには、そ

の不均衡を調整する役割が期待される。

1. CRC による情報提供の重要性

CRC が臨床試験に関する情報提供を行うことは、試験参加に関して患者の自己決定支援のために重要であるが、試験に参加している患者の安全確保と試験の正確性を担保するためにも重要となる。

a. 患者に情報提供の大切さを理解してもらう

患者の安全確保のためには、患者から適切な時期に正しく情報提供してもらうこと、副作用の対処法を理解してもらうことが必要である。臨床試験では適格基準を満たした患者を登録しなければならない。検査データなど測定できる基準はよいが、既往症・合併症・自覚症状などは本人からの情報提供が元になる。本人が使用している薬をすべて教えてもらわなければ、併用禁止薬を服用しているかどうか確認ができない。漢方薬・健康食品・サプリメントなどには相互作用のある場合もあり、情報把握が必要である。

こうした情報を正しく提供してもらわないと不適格な患者を登録してしまう可能性がある。また、思わぬ有害事象が発現するなど患者に不利益となる恐れもある。

試験開始後は、自覚症状の発現時期や程度を正しく伝えてもらうことは、データ収集上、必要であるばかりではなく、早期に適切な処置を行うことにつながる。さらに試験中に行う検査の意味や採血条件(空腹時・内服前か後か、など)も患者に十分理解してもらっておかなければ適切なデータを得るための協力が得られない。

情報提供は試験の正確性を担保するとともに患者の安全を守ることになり CRC は患者と信頼関係を築くことが求められる。

b. 有害事象が発現した際の対応と報告の重要性を知ってもらう

試験に参加した場合、試験中に起こりうる有害事象について、いつ頃起こるか、どのようなことが起こるか、起こったらどうすればよいかをあらかじめ情報提供しておけば、患者自身が慌てることなく対処できる。こうした有害事象に関する情報は、患者が最も知りたいことの1つであり、説明文書も重要な情報提供のツールとなる。

抗悪性腫瘍薬の臨床試験では、必ずといっていいほど有害事象が発現するため、つらくて途中で試験参加をやめたいときも、そのことを伝えることができる関係でなければならない。

しかし、臨床試験への参加を、患者からやめるとはいいだしにくいこともある。一方で試験治療を続けたいために、自覚症状について患者が報告しないこともあるかもしれない。CRCには言葉によるコミュニケーションだけでなく、態度や表情などから読みとれる能力も必要である。

2. コミュニケーションの6つのポイント

患者とのコミュニケーションには大きく分けて6つのポイントがある。

a. オープニング

初対面の時には、まず、あいさつと自己紹介を行う。CRCが何をする人かを説明してもよい。同席者については、できれば患者との関係を確認するとよい。

また、いきなり試験の説明に入らずに、相手がリラックスして説明を聞くことができるよう環境を整えることも大切である。

b. 共感的態度

良好なコミュニケーションをとるためには、話しやすいように目線や態度などにも注意が必要である。大げさすぎない笑顔で対応するなど表情にも気を配るとよい。

相手と目線の高さを合わせるように、立ったままではなく座って話をするにより、落ち着いて話をすることができ、重要な情報を収集することも少なくない。

会話中には、適度に相手の目を見る、うなづく・相槌を打つ・会話を促すことも大切である。それにより、相手が自分の話を聞いてもらっていると感ずることができ、また、事実を確認する言葉として「…と医師にいわれたのですね」など相手の言葉を繰り返す、「つらかったんですね」などの言葉をかけるとよい。

患者の話は、できる限り、妨げずに聞くほうがよいが、訴えが長い場合や横道にそれている場合は、話をまとめるなどして時間内に必要な情報を得るようにする能力も必要である。

c. 傾聴

がん臨床試験に参加する患者は自分自身の病気以外にもさまざまな不安を抱えていることが多い。

患者のさまざまな思い(病気に対する考え・家族への思い)や不安(治療に対する考え、生活設計、仕事などを含めた経済的問題)に耳を傾けることはよりよいコミュニケーションにつながり、重要な情報を得られることもある。患者の家族の会話も同様である。患者本人が、有害事象について過小に報告することもあるので身近にいる家族からの情報収集が有効な場合もある。

ただし、診察や検査の予約時間や自分自身の業務を考慮すること、適切な部署を紹介するなど自分の立場を考えた節度ある対応が必要である。

d. 説明・対応の方法

説明や対応は、時と場所を選ぶことが大切である。

患者自身に説明を聞く余裕がない場合もあり、CRCの状況にもよるが、その場で十分に時間がとれないときは、患者にその旨を伝え、相談したうえで、日時を改めて設定することが望まれる。

説明場所は、可能であれば診察室から少し離れた個室などが望ましい。

相手を尊重するという意味で原則として丁寧な言葉遣いを心がけるほうがよい。

説明に際して、専門用語はわかりやすく言い換えるなどが必要であるが、患者によっては治療歴が長いなどの理由で言い換える必要がないこともある。病歴や背景についてはカルテなどからわかる範囲で調べてもよい。

試験の内容についてはCRCが自分の言葉で説明できるよう、不明確な点は事前に医師などに確認しておきたい。なお、病状に関する説明などは、医師に直接説明してもらうように促す、あるいは診察に同席して、説明を依頼できるように配慮を行う。

e. 患者教育

患者が試験内容について理解できていない部分を確認、補足説明し、不安なところを補うことが求められる。そのうえで選択肢を示し、自己決定ができるようサポートを行う。

試験の内容など大まかなこと(スケジュール、点滴か内服か、入院か外来通院の

みか)は必ず理解してもらう必要がある。試験治療も、外来通院で行うことが多くなっているため、日常生活上の注意や有害事象が発現時の対処法など自己管理できるように情報提供が必要となる。

外来診察時は時間に限りがあるので、必要なことは要領よく医師に伝えることができるよう、CRCが診察前に情報を整理するとよい。

f. クロージング

最後に話忘れたことはないか、他に聞きたいことはないか、一言確認するとよい。また、次回の来院予定なども確認しておくとうい。

CRCにとってコミュニケーションスキルは十分な情報を得るためにも、提供するためにもとても重要である。

特にがんという疾患の特殊性から、患者やその家族の心情にも配慮し、表情や話しかたなどにも十分に気をつけて対応することが望まれる。

がん臨床試験に参加する患者にとって、CRCは医師や看護師より身近でコミュニケーションをとりやすい心強い存在であることが多い。また、そうした存在でなければならぬ。

けれども、CRCがすべて一人で対応するのではなく、必要に応じて、該当する部署のスタッフに対応を依頼、引き継ぎを行うとうい。

3. がん臨床試験におけるインフォームドコンセント

インフォームドコンセントとは、治療や検査などの医療行為について、患者本人が十分な説明を受け、理解し、納得したうえで同意することである。インフォームドコンセントに基づいて医療を受けることは、患者の自己決定権、すなわち基本的人権の1つとして保証されているものであり、臨床試験に限らず日常診療においても大切なことである。これは医療行為が少なからず侵襲を伴うからであるが、それに加えて臨床試験では、「治療効果や安全性について不確実性(uncertainty)が高く、リスクが大きい」などの理由により、日常診療の場合以上に詳細な説明が必要とされ、患者自身による自発的同意が臨床試験へ参加する際の必須条件となる。

本項では、がん臨床試験のインフォームドコンセントにおける留意事項について述べる。

Evaluating the 21-gene assay Recurrence Score[®] as a predictor of clinical response to 24 weeks of neoadjuvant exemestane in estrogen receptor-positive breast cancer

Takayuki Ueno · Norikazu Masuda · Takeharu Yamanaka · Shigehira Saji · Katsumasa Kuroi · Nobuaki Sato · Hiroyuki Takei · Yutaka Yamamoto · Shinji Ohno · Hiroko Yamashita · Kazufumi Hisamatsu · Kenjiro Aogi · Hiroji Iwata · Hironobu Sasano · Masakazu Toi

Received: 19 June 2013 / Accepted: 19 August 2013 / Published online: 8 October 2013
© Japan Society of Clinical Oncology 2013

Abstract

Background The aim of this study was to investigate the association between the results of the Recurrence Score (RS) assay and the clinical response to neoadjuvant endocrine therapy in postmenopausal women with breast cancer.

Methods Core biopsy samples at baseline and post-treatment surgical samples were obtained from 80 and 77 of 116 patients, respectively, enrolled in the multicenter prospective study of neoadjuvant exemestane therapy (JFMC34-0601). The 21-gene assay was performed after appropriate manual microdissection. The estrogen receptor

(ER), progesterone receptor, HER2 and Ki-67 were assayed by immunohistochemistry at a central laboratory. Clinical response was assessed based on the RECIST (Response Evaluation Criteria In Solid Tumors) guideline. **Results** Sixty-four core biopsy samples and 52 resection samples met the RS quality requirements. The clinical response rate in those patients with a low RS result (low RS group; 19/32, 59.4 %) was significantly higher than that in those patients with a high RS result (high RS group; 3/15, 20.0 %) ($P = 0.015$) and similar to that in patients with an intermediate RS result (intermediate RS group; 10/17, 58.8 %). The rates of breast-conserving surgery (BCS) were 90.6 % (29/32) in the low RS group, 76.5 % (13/17) in the intermediate RS group and 46.7 % (7/15) in the high RS group. The odds ratio for BCS adjusted for continuous

Takayuki Ueno and Norikazu Masuda have contributed equally to this study.

T. Ueno · M. Toi (✉)
Department of Breast Surgery, Graduate School of Medicine,
Kyoto University, 54 Kawara-cho, Shogoin, Sakyo-ku,
Kyoto 606-8507, Japan
e-mail: toi@kuhp.kyoto-u.ac.jp

N. Masuda
National Hospital Organization Osaka National Hospital,
Osaka, Japan

T. Yamanaka · S. Ohno
National Kyushu Cancer Center, Fukuoka, Japan

S. Saji
Department of Target Therapy Oncology, Kyoto University
Graduate School of Medicine, Kyoto, Japan

S. Saji · K. Kuroi
Tokyo Metropolitan Cancer and Infectious Diseases Center
Komagome Hospital, Tokyo, Japan

N. Sato
Niigata Cancer Center Hospital, Niigata, Japan

H. Takei
Division of Breast Surgery, Saitama Cancer Center, Saitama,
Japan

Y. Yamamoto
Department of Breast and Endocrine Surgery, Kumamoto
University, Kumamoto, Japan

H. Yamashita
Nagoya City University Graduate School of Medical Sciences,
Nagoya, Japan

K. Hisamatsu
Hiroshima City Asa Hospital, Hiroshima, Japan

K. Aogi
National Hospital Organization Shikoku Cancer Center, Ehime,
Japan

H. Iwata
Aichi Cancer Center Hospital, Nagoya, Japan

H. Sasano
Tohoku University School of Medicine, Sendai, Japan

baseline Ki-67 was 0.114 [95 % confidence interval (CI) 0.014–0.721; $P = 0.028$] between the high and low RS groups. RS values in pre-treatment samples were highly correlated with those in post-treatment samples (Spearman correlation coefficient 0.745, 95 % CI 0.592–0.846).

Conclusion Our results demonstrate the predictive value of the RS for clinical response to neoadjuvant exemestane therapy in postmenopausal women with ER-positive breast cancer.

Keywords Recurrence Score · Neoadjuvant endocrine therapy · Ki-67 · Clinical response · Breast-conserving surgery rate

Introduction

There are several potential advantages to neoadjuvant therapy of breast cancer in terms of improving outcomes in women with operable and inoperable early-stage disease [1, 2]. Both neoadjuvant chemotherapy and endocrine therapy have been shown to enable less extensive resection and improve rates of breast-conserving surgery (BCS) [3–6]. The ACOSOG Z1031 trial, which compared three aromatase inhibitors (AIs) in neoadjuvant settings, showed that 51 % (81/159) of the patients who were designated candidates for mastectomy experienced downstaging to BCS [7]. Neoadjuvant endocrine therapy is now an acceptable option for postmenopausal patients with endocrine-responsive disease [8].

Despite the use of standard biomarkers, the considerable heterogeneity of response to therapy still represents a challenge to clinicians in terms of choosing the most suitable neoadjuvant therapy. As such, tools to improve the identification of those patients who will respond to therapy would represent a major clinical advance. Although the Ki-67 labeling index (LI) shows some consistency in predicting response to chemotherapy, its ability to predict response to neoadjuvant endocrine therapy is controversial [9, 10].

We previously reported results from a neoadjuvant exemestane study in postmenopausal women [11]. In that study, the target response rate was 51 % (59/116), and 40 (77 %) of 59 patients who would have required mastectomy were converted to BCS. Neither baseline Ki-67 LI nor changes in Ki-67 LI were associated with clinical response in the study.

The *Oncotype DX*[®] assay (Genomic Health, Redwood City, CA) has been shown to be able assess recurrence risk in women with hormone receptor-positive (HR+), lymph node-negative or -positive, early stage breast cancer who are treated with adjuvant endocrine therapy [12–15]. It has also been shown to predict the likelihood of benefit from

adjuvant chemotherapy [12, 16]. Accordingly, the assay is included in clinical guidelines for use in patients with HR+ lymph node-negative disease; however, its applicability to HR+ postmenopausal women with lymph node positive disease is considered controversial, pending results of the RxPONDER trial [8, 17–19]. Additionally, studies in the neoadjuvant setting have shown that the test can be used to predict the response to chemotherapy [20, 21]. More recently, a study suggested that the Recurrence Score (RS) value may predict responses to neoadjuvant endocrine therapy with either tamoxifen or anastrozole [22]. The *Oncotype DX* assay may improve the clinician's ability to discriminate between clinically similar tumors based on the tumor's underlying biology. Consequently, the aim of this study was to investigate the clinical usefulness of the RS assay results in the prediction of response to neoadjuvant endocrine therapy.

Methods

Study design

This was a prospectively designed study using archived tumor tissues from the previously conducted JFMC34-0601 study. The primary objective was to assess the association between the results of the RS assay at baseline and clinical response, by comparing the response rates between patients with a low RS result (<18; low RS group) and those with a high RS result (≥ 31 ; high RS group). Secondary objectives included assessment of the associations of continuous baseline RS, quantitative estrogen receptor (ER) by reverse transcriptase (RT)-PCR and Ki-67 with clinical response and with BCS, as well as associations of changes from baseline to post-treatment values of these markers with clinical response. The study protocol was approved by the Ethics Committee of each participating institution. Informed consent was obtained from all patients. The study was performed in accordance with the Helsinki Declaration.

Patient cohort and tumor samples

Eligibility criteria for the parent JFMC34-0601 study included age 55–75 years, ER+ and stage II or IIIa invasive breast cancer (T2-3, N0-2, M0). Patients were confirmed positive for ER or progesterone receptor (PgR) by immunohistochemistry (≥ 10 % nuclear staining). The study treatment was 25 mg/day exemestane for 16 weeks, with a possible 8-week extension based on the assessment of clinical response. Patients with progressive disease (PD) were withdrawn from the study. At week 24, patients underwent surgery, except those with PD, who had the option of selecting another treatment approach.

Clinical outcomes measures

Clinical response was assessed by comparing the longest diameter of the target lesions with the baseline measurement, based on the Response Evaluation Criteria in Solid Tumors (RECIST) guideline version 1.0, by caliper measurement of palpable lesions and ultrasound as previously described [11]. Briefly, complete response (CR) was defined by the disappearance of all target lesions; partial response (PR) by at least a 30 % decrease in the sum of diameters of the target lesions; PD by at least a 20 % increase in the sum of diameters of the target lesions; stable disease (SD) by neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD.

Biomarker assessments

The *Oncotype DX*[®] 21-gene assay was performed on core biopsy and resection samples by Genomic Health [14].

Immunohistochemistry assays of Ki-67, ER and PgR were performed at one central location and the results assessed by three independent pathologists as described previously [11]. In brief, immunohistochemistry staining was performed using a Histofine kit (Nichirei, Tokyo, Japan). Ki-67 was stained using the following antibody dilution: 1:100 (Dako, Glostrup, Denmark), and the Ki-67 LI was obtained by counting 500–1,000 tumor cells at the sites of hot spots. Ki-67 groups were defined post hoc as <10, 10–30 and >30 %, respectively. ER and PgR immunoreactivity were scored according to Allred's procedure.

Expression of HER2 was determined by the HercepTest (Dako, Glostrup, Denmark). Positive HER2 status was defined as either 3+ or 2+ with confirmed c-erbB2 gene amplification by the fluorescence in situ hybridization (FISH) test.

Statistical analyses

Analyses of baseline markers included all patients with an evaluable RT-PCR result from core biopsies. Analyses of changes from baseline to post-treatment markers included the subset of patients with results from both core biopsies and surgical resections. Changes in continuous markers were defined as “post-treatment value–pre-treatment value”. In the primary analysis, the rates of clinical response were compared between the high and low baseline RS groups using Fisher's exact test. Logistic regression models were fit to both clinical response and surgery type. Odds ratio (OR) estimates are presented with Wald *p* values and 95 % confidence intervals (CIs). All *P* values are two-sided. In exploratory analyses, the Spearman rank correlation coefficient (and associated 95 % CI) was

calculated for the baseline continuous RS and either the post-treatment RS or baseline continuous Ki-67 as determined by immunohistochemistry. A paired *t* test was applied to compare the baseline and post-treatment RS values. A two-sample *t* test was used to compare the percentage reduction in tumor size between the high and low RS groups. Fisher's exact test was used to compare the conversion rate from mastectomy to BCS among risk groups.

Results

A total of 116 patients were enrolled in JFMC34-0601 between March 2006 and December 2007, of whom 102 completed 24 weeks of neoadjuvant exemestane treatment [11]. Core biopsy and resection samples were obtained for 80 (69 %) and 77 (66 %) patients, respectively. Of the 157 samples sent for *Oncotype DX* testing, two were deemed ineligible based on the blinded Genomic Health pathology review, insufficient RNA (<375 ng) was extracted from 18 samples (15 core biopsy and 3 resection samples), and standard quality metrics were not met for eight samples (all resections). This left 64 core biopsy samples, of which 52 had matching resection samples with evaluable RT-PCR results.

Baseline characteristics and clinical outcomes for the 64 patients are shown in Table 1. Forty-nine (76.6 %) patients had BCS, and 32 patients (50 %) had been candidates for BCS before the treatment. Four patients refused surgery after exemestane therapy and are treated as not BCS patients.

In the primary analysis, the clinical response rate in the low RS group (19/32, 59.4 %) was significantly higher than that in the high RS group (3/15, 20.0 %) ($P = 0.015$) (Table 2). The clinical response rate in the intermediate risk group (10/17, 58.8 %) was similar to that in the low risk group. Logistic regression revealed that the OR for clinical response between the intermediate and low RS groups was 0.977 (95 % CI 0.296–3.233, $P = 0.970$) and that the OR between the high and low RS groups was 0.171 (95 % CI 0.040–0.728, $P = 0.017$). In an exploratory analysis, the percentage reduction in tumor size determined by ultrasound was compared between the low and high RS groups. Patients in the low RS group showed an average reduction in tumor size of 31.8 % while those in the high RS group showed an average reduction of 12.5 %; this difference was significant between the groups ($P = 0.045$). The average reduction (27.6 %) in patients in the intermediate risk group was similar to that in the low risk group.

When treated as a continuous variable, the baseline RS Score was significantly associated with clinical response in a logistic regression analysis ($P = 0.042$; Table 3). There

Table 1 Baseline patient characteristics and clinical outcomes ($n = 64$)

Feature	n (%)
Age (years)	
55–64	34 (53.1)
65–74	25 (39.1)
75–77	5 (7.8)
Tumor stage at baseline	
T2	62 (96.9)
T3	2 (3.1)
Stage	
IIA	47 (73.4)
IIB	15 (23.4)
IIIA	2 (3.1)
ER by IHC (Allred score)	
4	1 (1.6)
5	3 (4.7)
6	5 (7.8)
7	14 (21.9)
8	41 (64.1)
ER status by RT-PCR	
ER– ($\leq 6.5 C_T$)	1 (1.5)
ER+ ($> 6.5 C_T$)	63 (98.4)
PgR by IHC (Allred score)	
0	4 (6.25)
4	7 (10.94)
5	4 (6.25)
6	8 (12.5)
8	12 (18.75)
NE	10 (15.63)
PgR status by RT-PCR	
PgR– ($\leq 5.5 C_T$)	14 (21.9)
PgR+ ($> 5.5 C_T$)	50 (78.1)
HER2 by IHC/FISH	
Negative	50 (78.1)
Positive	2 (3.1)
Unknown	12 (18.8)
RS risk group	
Low (< 18)	32 (50.0)
Intermediate (18–30)	17 (26.6)
High (≥ 31)	15 (23.4)
Ki-67 by IHC (%)	
< 10	28 (43.8)
10–30	23 (35.9)
> 30	13 (20.3)
Clinical response	
Complete response (CR)	0
Partial response (PR)	32 (50.0)
Stable disease (SD)	24 (37.5)
Progressive disease (PD)	5 (7.8)
NE	3 (4.7)

Table 1 continued

Feature	n (%)
Surgery type	
Breast-conserving	49 (76.6)
Mastectomy	11 (17.2)
No surgery	4 (6.3)

ER estrogen receptor, IHC immunohistochemistry, RT reverse transcriptase, PgR progesterone receptor, NE not evaluable, FISH fluorescence in situ hybridization, C_T cycling threshold score, RS recurrence Score

was a trend between continuous baseline ER as determined by RT-PCR and clinical response ($P = 0.076$). Continuous baseline Ki-67 by IHC was not associated with clinical response ($P = 0.273$).

The associations between changes from baseline to post-treatment values of continuous markers and clinical response were examined in logistic regression analyses. Changes in the RS, ER as determined by RT-PCR, and Ki-67 as determined by IHC were not associated with clinical response ($P = 0.240, 0.343$ and 0.629 , respectively).

Analysis of the RS categories and BCS is shown in Table 2. The OR for BCS between the intermediate and low RS groups was 0.336 (95 % CI 0.066–1.722, $P = 0.19$) and that between the high and low RS groups was 0.091 (95 % CI 0.019–0.432, $P = 0.003$). The logistic regression analyses of continuous baseline RS, ER by RT-PCR and Ki-67 by IHC with BCS are shown in Table 3. The continuous baseline RS was significantly associated with BCS in both the unadjusted ($p = 0.001$) and covariate-adjusted (for tumor size and PgR) ($P = 0.004$) analyses. The continuous baseline ER by RT-PCR was also significantly associated with BCS in both the unadjusted ($P = 0.001$) and covariate-adjusted ($P = 0.023$) analyses. Continuous baseline Ki-67 by IHC was significantly associated with BCS in the unadjusted analysis ($P = 0.024$) but lost its significance when adjusted for tumor size and PgR ($P = 0.060$). When both the continuous RS values and continuous Ki-67 were included in the logistic regression model for BCS, the RS retained its statistical significance ($P = 0.012$) whereas Ki-67 did not ($P = 0.868$). The conversion rate from mastectomy planned at baseline to BCS performed after the treatment was 88 % (15/17) in the low RS group, 70 % (7/10) in the intermediate RS group and 20 % (1/5) in the high RS group. The rate was significantly different among groups ($P = 0.010$).

The associations between RS and Ki-67, and their respective and joint associations with BCS were examined in exploratory analyses. Figure 1a shows a scatterplot of baseline Ki-67 as determined by IHC versus the baseline RS results. The Spearman correlation coefficient was 0.672 (95 % CI 0.506–0.785). All patients with PD had a high RS

Table 2 Clinical response and breast-conserving surgery according to categorical baseline Recurrence Score

RS risk group	Clinical response		
	Proportion (response rate) ^a (%)	Odds ratio (95 % CI)	P value
Low (RS <18)	19/32 (59.4)	1	n/a
Intermediate (RS 18–30)	10/17 (58.8)	0.977 (0.296, 3.233)	0.970
High (RS ≥31)	3/15 (20.0)	0.171 (0.040, 0.728)	0.017
RS risk group	Breast-conserving surgery		
	Proportion (BCS rate) (%)	Odds ratio (95 % CI)	P value
Low (RS <18)	29/32 (90.6)	1	n/a
Intermediate (RS 18–30)	13/17 (76.5)	0.336 (0.066, 1.722)	0.19
High (RS ≥31)	7/15 (46.7)	0.091 (0.019, 0.432)	0.003

Data are presented as the number of patients with the percentage in parenthesis

CI confidence interval, BCS breast-conserving surgery, n/a not available

^a Primary analysis: $P = 0.015$ by Fisher's exact test for comparison of clinical response rates between the low and high RS groups

Table 3 Continuous baseline Recurrence Score and estrogen receptor by reverse transcriptase-PCR and Ki-67 by immunohistochemistry and clinical response and breast-conserving surgery

Endpoint/analysis	Continuous marker					
	RS (50 units)		ER by RT-PCR (log ₂ increase)		Ki-67 by IHC (%)	
	Odds ratio (95 % CI)	P value	Odds ratio (95 % CI)	P value	Odds ratio (95 % CI)	P value
Clinical response/unadjusted	0.205 (0.044, 0.946)	0.042	1.436 (0.963, 2.141)	0.076	0.981 (0.948, 1.015)	0.273
BCS/unadjusted	0.055 (0.009, 0.323)	0.001	1.786 (1.150, 2.774)	0.001	0.957 (0.921, 0.994)	0.024
BCS/covariate-adjusted ^a	0.016 (<0.001, 0.259)	0.004	1.881 (1.090, 3.245)	0.023	0.953 (0.907, 1.002)	0.060

RT reverse transcriptase

^a Adjusted for tumor size and PgR Allred score, which were significantly associated with BCS in the univariable analyses

(range 32–73) while three of five PD patients had an intermediate Ki-67 LI (Fig. 1a).

No statistically significant difference was observed between baseline and post-treatment RS values ($P = 0.484$). A scatterplot is shown in Fig. 1b. The Spearman correlation analysis showed a high correlation (correlation coefficient 0.745, 95 % CI 0.592–0.846).

Discussion

In this study, we demonstrated the predictive value of the RS results for response to neoadjuvant endocrine therapy. Among our patient cohort, those with low scores showed a better response to neoadjuvant endocrine therapy than those with high scores. Since patients with high RS results have been shown to benefit from chemotherapy, the 21-gene assay may provide additional information that could facilitate the selection of neoadjuvant treatment with endocrine therapy for cancer

patients with a low RS and chemotherapy for those with a high RS.

ER Allred scores have been reported to correlate with response rates to neoadjuvant letrozole or tamoxifen. The P024 trial of neoadjuvant letrozole or tamoxifen showed that tumors with low ER Allred scores still responded to letrozole [23]. Conversely, some tumors with higher ER levels did not respond to endocrine therapy [23, 24]. Gene expression-based profiles categorize HR+, HER2– breast cancers into two subtypes: luminal-A and -B [25]. However, the classification, which is based on PAM50, has been reported not to relate to clinical response or the likelihood of BCS after neoadjuvant AI treatment [7].

In our study, the RS was the only predictive factor for clinical responses to neoadjuvant endocrine therapy and the most potent predictive factor for BCS in the covariate-adjusted analysis. These results are consistent with those from other studies which suggest that a low RS can predict benefit from endocrine therapy [22, 24]. The study by Kim et al. [24] compared the outcomes of the tamoxifen and

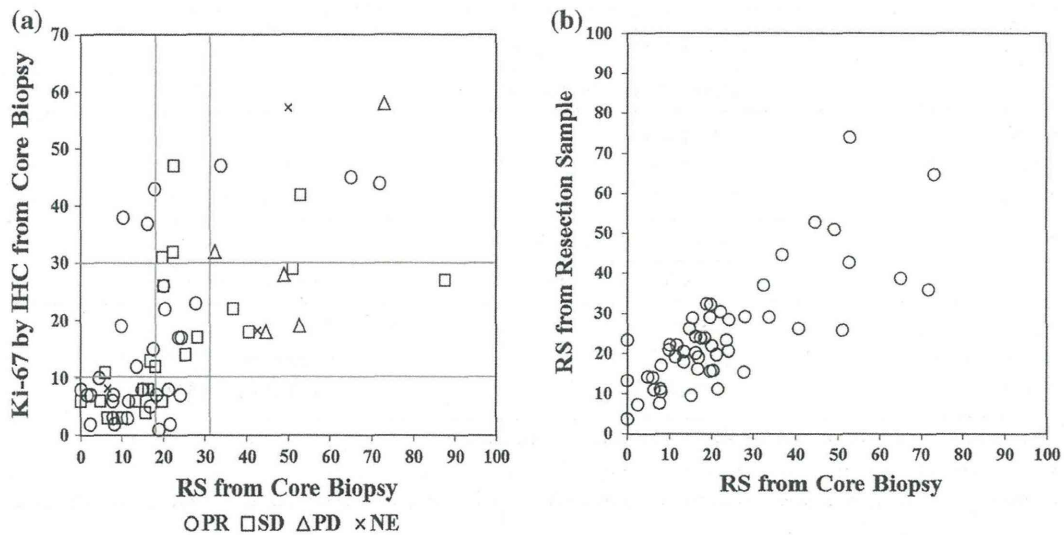


Fig. 1 a Scatterplot of the baseline Recurrence score (RS) and baseline Ki-67, with the Spearman correlation coefficient. The Spearman correlation coefficient between the baseline RS and baseline Ki-67 was 0.672 [95 % confidence interval (CI) 0.506–0.785]. None of five patients with tumor progression was in the low or intermediate RS groups. **b** Scatterplot of the baseline RS

and post-treatment RS, with the Spearman correlation coefficient. The baseline RS was highly correlated with RS in the post-treatment samples (Spearman correlation coefficient 0.745, 95 % CI 0.592–0.846). *PR* Partial response, *SD* stable disease, *PD* progressive disease, *NE* Not evaluable

placebo arms of the NSABP B14 trial and demonstrated that higher levels of quantitative ER expression, as determined by RT-PCR, correlated with a greater benefit from adjuvant tamoxifen, as measured by distant recurrence.

Our results indicate that the values of the RS before and after endocrine therapy were highly correlated. Since a number of studies have suggested that post-treatment biomarkers such as Ki-67 LI and ER have better prognostic values than pre-treatment biomarkers, post-treatment biomarkers are receiving increasing interest in clinical trials as a tool for patient stratification [26–28]. Dowsett et al. [26] reported the results of an unplanned, exploratory investigation of the relationship between post-treatment Ki-67 (2 weeks) and recurrence-free survival (RFS) using archived tumors from the IMPACT study. Their results indicate that post-treatment Ki-67, larger baseline tumor size and post-treatment ER level are significantly correlated with DFS. Ellis et al. [27] analyzed the ability of post-treatment Ki-67 and other factors (tumor size, grade, nodal status, and post-treatment ER expression) to predict RFS and breast cancer-specific survival using archived tumors from the P024 study. Another interesting study (ACOSOG Z1031, Cohort B) has been conducted to determine whether patients with a high Ki-67 value after 2 weeks of neoadjuvant AI treatment show a higher than expected pathologic CR rate to neoadjuvant chemotherapy than would be typically observed for those patients with unselected ER-rich tumors. The results will tell us whether an assessment of

Ki-67 2 weeks after neoadjuvant AI treatment will be useful for the identification of a chemotherapy-sensitive subgroup of ER+ tumors. However, even if this is the case, intervention of a 2-week AI treatment and re-biopsy are necessary. Although further investigations are needed, the comparative stability of the RS would improve the overall decision-making process regarding the complete treatment before the initiation of treatment.

The main limitation of this was its small sample size. The availability of tumor samples from the parent study was limited and recovery of mRNA was not uniformly adequate. Further investigation in larger prospective studies would better define candidates for neoadjuvant endocrine therapy. Another limitation was the absence of any assessment of lymph node response. Although nodal response is clinically relevant, one of the major purposes of neoadjuvant endocrine therapy is improvement in surgical outcome. That said, however, the clinical response at the primary site and the BCS rate are also of clinical importance for the assessment of the effect of neoadjuvant endocrine therapy.

In conclusion, this study showed that RS results have predictive value for the clinical response to neoadjuvant exemestane therapy. The 21-gene assay would appear to be a promising tool for providing useful information to guide the clinician in choosing neoadjuvant treatment for systemic therapy, with neoadjuvant endocrine treatment for patients with low RS disease and neoadjuvant chemotherapy treatment for patients with high RS disease.

Acknowledgments This trial was supported by The Japanese Foundation for Multidisciplinary Treatment of Cancer. We thank all the members of JFMC34-0601: G. Wakabayashi, Iwate Medical University; H. Bando, Tsukuba University; S. Nakamura, Showa University Hospital; R. Nishimura, Kumamoto City Hospital; S. Amano, Nihon University Itabashi Hospital; T. Ohmura, Sapporo Medical University; Y. Yanagida, Gunma Prefectural Cancer Center; T. Saeki, Saitama Medical University International Medical Center; K. Kojima, Juntendo University Hospital; T. Sawada, Showa University Hospital; H. Ogata, Toho University Omori Medical Center; H. Yasuda, International Medical Center of Japan; S. Takahashi, The Cancer Institute Hospital of JFCR; M. Takami, Tokyo Metropolitan Fuchu Hospital; T. Nishi, Mitsui Memorial Hospital; A. Chiba, Kanagawa Cancer Center; Y. Tokuda, Tokai University; K. Ito, Shinshu University Hospital; T. Utsumi, Fujita Health University; K. Anan, Kitakyushu Municipal Medical Center. We thank Dr. Shigetoyo Saji, Dr. Yoshitaka Furuta, Ms Minako Nakashima and Ms. Kikuko Fujita for their support.

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

References

- Fisher B, Brown A, Mamounas E et al (1997) Effect of preoperative chemotherapy on local-regional disease in women with operable breast cancer: findings from national surgical adjuvant breast and bowel project B-18. *J Clin Oncol* 15:2483–2493
- Chia S, Swain SM, Byrd DR et al (2008) Locally advanced and inflammatory breast cancer. *J Clin Oncol* 26:786–790
- Makris A, Powles TJ, Ashley SE et al (1998) A reduction in the requirements for mastectomy in a randomized trial of neoadjuvant chemoendocrine therapy in primary breast cancer. *Ann Oncol* 9:1179–1184
- Mauriac L, MacGrogan G, Avril A et al (1999) Neoadjuvant chemotherapy for operable breast carcinoma larger than 3 cm: a unicentre randomized trial with a 124-month median follow-up. Institut Bergonie Bordeaux Groupe Sein (IBBGS). *Ann Oncol* 10:47–52
- Mauri D, Pavlidis N, Ioannidis JP (2005) Neoadjuvant versus adjuvant systemic treatment in breast cancer: a meta-analysis. *J Natl Can Inst* 97:188–194
- Semiglazov VF, Semiglazov VV, Dashyan GA et al (2007) Phase 2 randomized trial of primary endocrine therapy versus chemotherapy in postmenopausal patients with estrogen receptor-positive breast cancer. *Cancer* 110:244–254
- Ellis MJ, Suman VJ, Hoog J et al (2011) Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2–3 breast cancer: clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype-ACOSOG Z1031. *J Clin Oncol* 29:2342–2349
- Goldhirsch A, Wood WC, Coates AS et al (2011) Strategies for subtypes—dealing with the diversity of breast cancer: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer 2011. *Ann Oncol* 22:1736–1747
- Chang J, Powles TJ, Allred DC et al (2000) Prediction of clinical outcome from primary tamoxifen by expression of biologic markers in breast cancer patients. *Clin Can Res* 6:616–621
- Miller WR, White S, Dixon JM et al (2006) Proliferation, steroid receptors and clinical/pathological response in breast cancer treated with letrozole. *Br J Can* 94:1051–1056
- Toi M, Saji S, Masuda N et al (2011) Ki-67 index changes, pathological response and clinical benefits in primary breast cancer patients treated with 24 weeks of aromatase inhibition. *Can Sci* 102:858–865
- Albain KS, Barlow WE, Shak S et al (2010) Prognostic and predictive value of the 21-gene recurrence score assay in postmenopausal women with node-positive, oestrogen-receptor-positive breast cancer on chemotherapy: a retrospective analysis of a randomised trial. *Lancet Oncol* 11:55–65
- Mamounas EP, Tang G, Fisher B et al (2010) Association between the 21-gene recurrence score assay and risk of loco-regional recurrence in node-negative, estrogen receptor-positive breast cancer: results from NSABP B-14 and NSABP B-20. *J Clin Oncol* 28:1677–1683
- Paik S, Shak S, Tang G et al (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *New Engl J Med* 351:2817–2826
- Tang G, Shak S, Paik S et al (2011) Comparison of the prognostic and predictive utilities of the 21-gene recurrence score assay and adjuvant! For women with node-negative, ER-positive breast cancer: results from NSABP B-14 and NSABP B-20. *Breast Can Res Treat* 127:133–142
- Paik S, Tang G, Shak S et al (2006) Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer. *J Clin Oncol* 24:3726–3734
- National Comprehensive Cancer Network (NCCN) (2013) Clinical practice guidelines in oncology (NCCN®). Available at: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site
- Aebi S, Davidson T, Gruber G et al (2011) Primary breast cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 22[Suppl 6: vi]:12–24
- Harris L, Fritsche H, Mennel R et al (2007) American society of clinical oncology 2007 update of recommendations for the use of tumor markers in breast cancer. *J Clin Oncol* 25:5287–5312
- Chang JC, Makris A, Gutierrez MC et al (2008) Gene expression patterns in formalin-fixed, paraffin-embedded core biopsies predict docetaxel chemosensitivity in breast cancer patients. *Breast Can Res Treat* 108:233–240
- Gianni L, Zambetti M, Clark K et al (2005) Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *J Clin Oncol* 23:7265–7277
- Akashi-Tanaka S, Shimizu C, Ando M et al (2009) 21-Gene expression profile assay on core needle biopsies predicts responses to neoadjuvant endocrine therapy in breast cancer patients. *Breast* 18:171–174
- Ellis MJ, Ma C (2007) Letrozole in the neoadjuvant setting: the P024 trial. *Breast Can Res Treat* 105[Suppl 1]:33–43
- Kim C, Tang G, Pogue-Geile KL et al (2011) Estrogen receptor (ESR1) mRNA expression and benefit from tamoxifen in the treatment and prevention of estrogen receptor-positive breast cancer. *J Clin Oncol* 29:4160–4167
- Sorlie T, Perou CM, Tibshirani R et al (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98:10869–10874
- Dowsett M, Smith IE, Ebbs SR et al (2007) Prognostic value of Ki-67 expression after short-term presurgical endocrine therapy for primary breast cancer. *J Natl Can Inst* 99:167–170
- Ellis MJ, Tao Y, Luo J et al (2008) Outcome prediction for estrogen receptor-positive breast cancer based on post neoadjuvant endocrine therapy tumor characteristics. *J Natl Can Inst* 100:1380–1388
- Chia YH, Ellis MJ, Ma CX (2010) Neoadjuvant endocrine therapy in primary breast cancer: indications and use as a research tool. *Br J Can* 103:759–764

Sarcopenia as a predictor of prognosis in patients following hepatectomy for hepatocellular carcinoma

N. Harimoto¹, K. Shirabe¹, Y.-I. Yamashita¹, T. Ikegami¹, T. Yoshizumi¹, Y. Soejima¹, T. Ikeda¹, Y. Maehara¹, A. Nishie² and T. Yamanaka³

¹Department of Surgery and Science and ²Department of Clinical Radiology, Graduate School of Medical Sciences, Kyushu University, Fukuoka, and

³Biostatistics Section, Research Centre for Innovative Oncology, National Cancer Hospital East, Chiba, Japan

Correspondence to: Dr N. Harimoto, Department of Surgery and Science, Graduate School of Medical Sciences, Kyushu University, 3-1-1 Maidashi, Higashi-ku, Fukuoka 812-8582, Japan (e-mail: nharimoto1@fukuoka.email.ne.jp)

Background: Sarcopenia was identified recently as a poor prognostic factor in patients with cancer. The present study investigated the effect of sarcopenia on short- and long-term outcomes following partial hepatectomy for hepatocellular carcinoma (HCC), and aimed to identify prognostic factors.

Methods: Data were collected retrospectively for all consecutive patients who underwent hepatectomy for HCC with curative intent between January 2004 and December 2009. Patients were assigned to one of two groups according to the presence or absence of sarcopenia, assessed by computed tomographic measurement of muscle mass at the level of the third lumbar vertebra. Clinicopathological, surgical outcome and long-term survival data were analysed.

Results: Sarcopenia was present in 75 (40.3 per cent) of 186 patients, and was significantly correlated with female sex, lower body mass index and liver dysfunction, as indicated by abnormal serum albumin levels and indocyanine green retention test at 15 min values. In patients with, and without sarcopenia, the 5-year overall survival rate was 71 and 83.7 per cent respectively, and the 5-year recurrence-free survival rate was 13 and 33.2 per cent respectively. Multivariable analysis revealed that reduced skeletal muscle mass was predictive of an unfavourable prognosis.

Conclusion: Sarcopenia was predictive of worse overall survival even when adjusted for other known predictors in patients with HCC after partial hepatectomy.

Paper accepted 8 July 2013

Published online in Wiley Online Library (www.bjs.co.uk). DOI: 10.1002/bjs.9258

Introduction

Hepatocellular carcinoma (HCC) is one of the most common malignancies in the world^{1,2}. As a consequence of advances in the diagnosis and management of HCC, major improvements in overall and disease-free survival rates for HCC after partial hepatectomy have been achieved. However, even when curative resection is performed, a considerable number of patients develop intrahepatic or extrahepatic recurrence^{3,4}. The prognostic assessment of patients with HCC after hepatic resection and recurrence is an important clinical issue in this population⁵⁻⁷. Both tumour- and host-related factors are related to clinical outcome, and general condition and liver function are important in this context. Unfortunately, it is difficult to evaluate the general condition of patients excluding liver function before hepatectomy. Conventional methods, such as the Child–Pugh classification, have been used

initially to determine the severity of cirrhosis and to select patients who might tolerate hepatic resection. However, these methods do not reflect the patient's general condition. The American Society of Anesthesiologists (ASA) grade was reported to predict the prognosis of HCC after hepatectomy⁸, but this classification is not always objective.

Recently, loss of skeletal muscle mass, termed sarcopenia, was identified as a poor prognostic factor for patients with pancreatic cancer, colorectal liver metastases, melanoma, liver cirrhosis and liver transplantation⁹⁻¹⁴. Sarcopenia is a syndrome characterized by progressive and generalized loss of skeletal muscle mass and strength, with a risk of adverse outcomes such as physical disability, poor quality of life and death^{15,16}. To date, there have been no reports on the relationship between sarcopenia and the prognosis of patients with HCC following hepatic resection.

A retrospective study was performed at the authors' institution to investigate the outcome of patients with sarcopenia who underwent hepatic resection for HCC. The outcome of these patients was compared with that of patients without sarcopenia undergoing hepatic resection during the same period.

Methods

All patients who underwent hepatic resection with curative intent as the initial treatment in the Department of Surgery II, Kyushu University Hospital, between January 2004 and December 2009 were enrolled in the study. Curative resection was defined as complete macroscopic removal of the tumour. All patients had preoperative computed tomography (CT). A transverse CT image at the third lumbar vertebra (L3) in the inferior direction was assessed from each scan. Skeletal muscle was identified and quantified by Hounsfield unit (HU) thresholds of -29 to $+150$ (water is defined as 0 HU, air as 1000 HU). Multiple muscles were quantified, including the psoas, erector spinae, quadratus lumborum, transversus abdominis, external and internal oblique abdominal muscle, and rectus abdominis muscle (Fig. 1). CT measurements were calibrated with water and air at fixed intervals. Cross-sectional areas (cm^2) of skeletal muscles in the L3 region were measured by manual outlining on the CT images, and checked by the radiologist. The cross-sectional areas were then normalized for height (cm^2/m^2).

Cut-off values for skeletal muscle associated with overall survival were defined as $43.75 \text{ cm}^2/\text{m}^2$ for men and $41.10 \text{ cm}^2/\text{m}^2$ for women¹⁰. Based on this cut-off, patients were assigned to one of two groups, depending on the presence or absence of sarcopenia. The clinicopathological

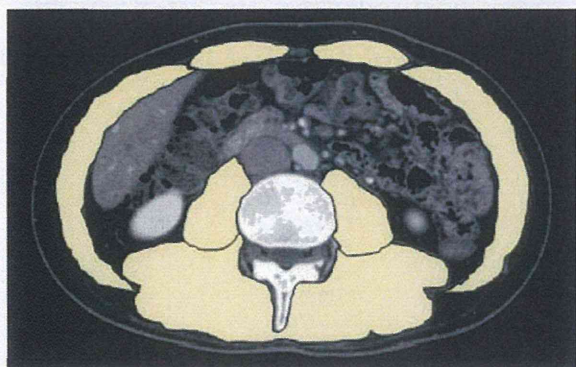


Fig. 1 Computed tomogram showing the area of skeletal muscle mass in the L3 region (highlighted yellow)

background and rates of overall and recurrence-free survival were compared between the two groups.

The prognostic factors were examined with respect to overall and recurrence-free survival on the basis of the following variables: sarcopenia (absence *versus* presence); skeletal muscle mass; age; sex (male *versus* female); body mass index (BMI); hepatitis B surface antigen (positive *versus* negative), hepatitis C virus antibody (positive *versus* negative); serum albumin level; serum total bilirubin level; serum aspartate aminotransferase level; platelet number; indocyanine green retention test at 15 min (ICGR15); Child–Pugh grade (A *versus* B); Model for End-Stage Liver Disease (MELD) score; histological liver cirrhosis (normal liver + chronic hepatitis *versus* liver fibrosis and liver cirrhosis); tumour size; tumour number (solitary *versus* multiple); tumour node metastasis (TNM) stage according to the Liver Cancer Study Group of Japan¹⁷ (I+II *versus* III+IV); tumour differentiation (well differentiated + moderately differentiated *versus* poorly differentiated); microvascular invasion (MVI) (absence *versus* presence); intrahepatic metastases (absence *versus* presence); serum α -fetoprotein level (AFP); des- γ -carboxyprothrombin (DCP) level; operative procedure (anatomical *versus* non-anatomical resection); duration of surgery; estimated blood loss; and postoperative complications (absence *versus* presence). Patients with diabetes were defined as those using an oral hypoglycaemic agent or insulin. The MELD score was calculated in accordance with a previous report¹⁸. Postoperative complications within 1 month after partial hepatectomy included liver failure, encephalopathy, gastrointestinal bleeding, intraperitoneal abscess, abdominal haemorrhage, bile leakage, pleural effusion, intractable ascites and wound infection. Complications were classified according to Clavien–Dindo¹⁹; grade III complications (those requiring surgical intervention) were considered to indicate the presence of a postoperative complication.

Surgical procedures

Details of surgical techniques and patient selection criteria have been reported previously⁷. Selection criteria for hepatic resection were: ascites not detected, or controllable by diuretics; serum total bilirubin level lower than 2.0 mg/ml ; and ICGR15 value below 40 per cent. The surgical approach included a J-shaped incision for routine abdominal access, hepatic dissection using an ultrasonic dissector with a coagulator (CUSA EXcel®; Integra, Plainsboro, New Jersey, USA), with systematic ligation of all sizable vessels, and close ultrasonographic guidance along the transection line. Cholecystectomy was performed

in all patients if applicable. An intraoperative bile leak test was performed routinely²⁰. Small bile leaks on the cut liver surface were repaired by Z-suturing with 6-0 polydioxanone (PDS II; Johnson and Johnson, Tokyo, Japan). Intraoperative vascular control was achieved with the Pringle manoeuvre²¹.

Follow-up strategy and recurrence pattern

After discharge, all patients were examined monthly for recurrence by ultrasonography and estimation of tumour markers, such as AFP and DCP, and by CT every 6 months. When recurrence was suspected, additional examinations such as hepatic angiography were performed. Recurrent

HCC was treated by repeat hepatectomy, ablation therapy and lipiodolization, as described previously²².

Histological assessment

All resected specimens were cut into serial 5–10-mm thick slices and fixed in 10 per cent formalin. After macroscopic examination, the slice with the greatest dimensions was trimmed for embedding in paraffin and cut into 4- μ m microscopic sections. The sections were stained with haematoxylin and eosin. Tumour differentiation, MVI, intrahepatic metastases and histological liver cirrhosis were assessed by the pathologist in accordance with the rules of the Liver Cancer Study Group of Japan¹⁷.

Table 1 Clinicopathological factors in patients with, and without sarcopenia

	Sarcopenia (n = 75)	No sarcopenia (n = 111)	P†
Age (years)	67(11)	66(10)	0.553
Sex ratio (M:F)	50:25	95:16	0.004‡
Skeletal muscle mass (cm ² /m ²)	37.8(3.7)	49.7(6.5)	<0.001
Body mass index (kg/m ²)	20.5(2.4)	24.0(2.8)	<0.001
Diabetes mellitus	22 (29)	35 (31.6)	0.999‡
Albumin (g/dl)	3.8(0.4)	4.0(0.4)	0.002
Total bilirubin (mg/dl)	0.9(0.4)	0.8(0.3)	0.096
Platelet count ($\times 10^4/\mu$ l)	15.5(7.5)	16.3(6.2)	0.454
ICGR15 (%)	15.7(8.2)	13.6(6.2)	0.049
Child–Pugh grade			0.190‡
A	68 (91)	107 (96.4)	
B	7 (9)	4 (3.6)	
MELD score	7.7(2.1)	7.9(1.8)	0.591
Hepatitis grade			0.652‡
None	11 (15)	13 (11.7)	
Mild	55 (73)	80 (72.1)	
Severe	9 (12)	18 (16.2)	
Liver cirrhosis			0.290‡
Normal liver + chronic hepatitis	32 (43)	55 (49.5)	
Liver fibrosis + liver cirrhosis	43 (57)	56 (50.5)	
Tumour size (cm)	4.0(3.2)	3.9(2.8)	0.770
No. of tumours			0.171‡
Solitary	52 (69)	88 (79.3)	
Multiple	23 (31)	23 (20.7)	
TNM stage			0.967‡
I	11 (15)	18 (16.2)	
II	38 (51)	57 (51.4)	
III	20 (27)	29 (26.1)	
IV	6 (8)	7 (6.3)	
Differentiation of HCC			0.690‡
Well	9 (12)	10 (9.0)	
Moderate	50 (67)	77 (69.4)	
Poor	16 (21)	24 (21.6)	
Microvascular invasion	24 (32)	37 (33.3)	0.890‡
Intrahepatic metastases	12 (16)	18 (16.2)	0.978‡
α -Fetoprotein level (ng/ml)	3459(18 300)	12 250(70 470)	0.297
DCP (munits/l)	4318(13 627)	2942(12 499)	0.480
Postoperative complications	24 (32)	56 (50.5)	0.613‡

Values are mean(s.d.) unless indicated otherwise: *values in parentheses are percentages. ICGR15, indocyanine green dye retention test at 15 min; MELD, Model for End-Stage Liver Disease; TNM, tumour node metastasis (stage defined by the Liver Cancer Study Group of Japan); HCC, hepatocellular carcinoma; DCP, des- γ -carboxyprothrombin. †Mann–Whitney *U* test, except ‡Fisher's exact test or χ^2 test.