

第2章 イレッサの有用性評価

第1節 イレッサの市販前の有効性評価

第1 抗ガン剤の有効性評価の基本的な指標…真のエンドポイント

1 はじめに

第1章で見たとおり，医薬品の有効性は科学的に証明された場合にのみ認められるのが原則であり，科学的な証明がなされない限り，当該医薬品は無効であると評価されなければならない。そして，科学的な証明は，適正にデザインされ，適切な真のエンドポイントを用いた比較臨床試験の結果によらなければならない。

これを抗ガン剤において見れば，抗ガン剤の有効性評価における真のエンドポイントは延命効果であり，全生存期間（overall survival）の延長である。

こうした点について，福島雅典証人は，以下のとおり述べている。

「通常，抗癌剤の有効性と言った場合には，これは延命効果のあり，なしを指します。…最終的に腫瘍が小さくなくてもそれが延命につながるかどうかは別の問題です。…単純に血圧を下げました，ああ，有効ですと，この薬を飲みましようと言ってたら，どういう副作用が出るか分からないし，実際に延命につながるかどうか，つまり，脳卒中や心筋梗塞や，そういうイベントを抑えるかどうか分からない。これを真のエンドポイント，真の評価ポイントと言います。ですから，抗癌剤の評価は，必ず延命できるかどうかで評価しないとイケないんです。単に，その薬を飲んだら，あるいは注射したら，腫瘍が小さくなりましたと。じゃ，1か月経ったら元の木阿弥で，そのあと

はリバウンドで、もっと大きくなってしまいましたでは、全く有効性があるとは言えないわけですから、ですから、真のエンドポイントである延命効果で見ない限りは、有効性については議論できないということです。」（西福島証人主尋問調書＝東甲L95p29, 30）

「厳密な比較臨床試験をしない限りは、有効性については議論できない。延命効果については議論できないということになります。」（同p30）

被告らは、腫瘍縮小効果や延命効果、QOL、個別症例報告などの様々な指標をごちゃ混ぜにして「総合判断」することにより有用性を肯定できるかに主張し、被告側証人である福岡証人も同様の証言をした部分もあるが、こうした被告らの主張が破綻していることは明白である。

なお、イレッサ承認当時の「平成3年『抗悪性腫瘍薬の臨床評価方法に関するガイドライン』について」（いわゆる旧ガイドライン、西乙D7＝東乙H7）についての検討は後に行うが、まず、何よりも抗ガン剤の有効性評価の真のエンドポイントは、本来、延命効果であることを強く認識する必要がある、そのことが旧ガイドラインの評価においても極めて重要な意義を持つのである。

以下、さらに具体的に検討する。

2 抗ガン剤の臨床試験におけるエンドポイント

(1) エンドポイントの種類

「臨床試験のエンドポイント」（「癌の臨床」49巻10号,2003/10西甲F30＝東G56）は、国立がんセンター研究所がん情報研究部・JCOG（日本臨床腫瘍研究グループ）データセンターに所属している山本精一郎氏が執筆したもので、ガンの臨床試験におけるエンドポイントについて一般的に解説している文献である。

臨床試験におけるエンドポイントとは、研究の目的ないしは指標という意

味で用いられる用語であり（西甲F 3 0 = 東G 5 6 p 1 2 1 1），それぞれの臨床試験における測定尺度（例えば，比較対照群を持たない第Ⅱ相試験における腫瘍反応割合）ないしは証明対象（例えば，比較対照群を持つ第Ⅲ相試験における生存期間の優越性，非劣性等）を指す。

エンドポイントには，「真のエンドポイント」と「サロゲートエンドポイント」があり，「真のエンドポイント」は，患者がどのように感じ，機能し，生存しているかを直接測定する臨床的に意味のあるエンドポイントであり，「サロゲートエンドポイント」は，この真のエンドポイントの代替となるものである（西甲F 3 0 = 東G 5 6 p 1 2 1 2）

また，別の観点からは，「ハードなエンドポイント」と「ソフトなエンドポイント」という種別もあり，「測定されたエンドポイントが客観的で誤差のない場合をハードなエンドポイントと呼び，奏効率や無再発生存期間のように測定者や測定日によって結果が変化するようなものをソフトなエンドポイントと呼ぶ」とされている（甲F 3 0 p 1 2 1 2， 1 2 1 3）

そして，臨床試験においては，「どの段階でもできるだけ真のエンドポイントか効果の証明されたサロゲート・エンドポイントを用いること，できるだけハードで一般的なエンドポイントを用いるべきである。」とされている（西甲F 3 0 = 東G 5 6 p 1 2 1 7）。

以下，個別に検討する。

（2） 生存期間

生存期間は，臨床試験登録日から死亡までの時間として定義されるtime-to-eventの一つである（西甲F 3 0 = 東G 5 6 p 1 2 1 3，ここでの生存期間は，「全生存期間—overall survival—と同義である。）。

この生存期間は真のエンドポイントであり（西甲F 3 0 = 東G 5 6 p 1 2 1 2），また，「研究者の解釈やマスク化していないことによるバイアスの

影響を受けないハードなエンドポイントといえる。」とされている（甲F30＝東G56 p1213）。

このため、「多くのがん第Ⅲ相臨床試験の真のエンドポイントは生存期間であり」とされ（西甲F30＝東G56 p1212），また，被告側証人である西條長宏証人ですら，「endpointとしてはもちろん第一に延命効果である。もし生存期間に差が出ない場合は，無病生存期間(disease free survival), QOL, cost benefitなどが検討されることがある。」（西甲F32＝東G58 p24）としている。

なお，抗ガン剤の第Ⅲ相臨床試験における生存期間の検定は，プラセボもしくは無治療・緩和療法群（BSC群）に対する優越性試験（イレッサにおけるISEL試験）と，既存の標準治療に対する非劣性試験（イレッサにおけるドセタキセルとの比較国内第Ⅲ相試験）があるところ，非劣性試験においては，非劣性限界，非劣性の許容範囲の設定（イレッサにおけるドセタキセルとの比較国内第Ⅲ相試験においては，非劣性限界としての95%信頼区間の上限値を1.25に設定するデザインをし，実際の試験では95%信頼区間の上限が1.4と上限値を上回ったため，イレッサのドセタキセルに対する非劣性は証明されなかった（西甲C5＝東E10－3））などの試験デザインの妥当性が慎重に吟味される必要がある（西甲F30＝東G56 p1213, 1214）。

（3） 腫瘍縮小効果

腫瘍縮小効果は，薬剤自体の効果を測る上では真のエンドポイントではあるが，生存を真のエンドポイントとした場合のサロゲート・エンドポイントとして優れているかどうかは疾患の種類に依存するとされる（西甲F30＝東G56 p1215）。

また，「腫瘍縮小効果の複雑な側面として，評価可能病変の事前特定，評

価のタイミング，画像による評価という方法論的な難しさが挙げられる。同じ奏効率でもCRとPRの割合が異なる場合，奏効期間が異なる場合，腫瘍が縮小した部位の違い，症状改善との関連，転移の程度と大きさをどう扱うかといった問題を含むソフトなエンドポイントである。」とされている（西甲F30＝東G56 p1215）。

福岡証人や西條証人は，腫瘍縮小と延命効果には相関関係があるとするが，腫瘍縮小は延命効果が認められるための必要条件ではあっても，腫瘍縮小が認められたからといって延命効果が認められるという意味での十分条件ではあり得ない。もし，十分条件としての意味も持つとすれば，生存期間をエンドポイントとした第Ⅲ相試験は必要ないこととなるはずであるが，現実には，そのようなものとなっていないことは明らかである。この点は，さらに後述する。

（４） QOL，症状改善等の患者の主観的なエンドポイント

ア QOL，症状改善等は，患者に対するアンケートを主体として調査したものであるため，患者の主観的なものであり，非常にソフトなエンドポイントである（西甲F30＝東G56 p1216，西丙E49の1＝東坪井証人反対尋問調書p79～81）。このため，「正しくそれを測定するためには，妥当性や再現性のある尺度を用いることが必要」である（西甲F30＝東G56 p1216）。

特に，QOL調査は，直接病状について聞く肺癌サブスケールの項目以外にも，「友人を身近に感じる」等の質問項目が多数取り入れられており（西丙H44＝東甲L157），抗ガン剤の治療効果以外の要素に大きな影響を受けることが避けられないものとなっている（西丙E49の1＝東坪井証人反対尋問調書p79。80）。

こうしたことから，西條証人は，「QOLは延命効果で差を検出しがた

い癌腫に対する第Ⅲ相研究でのendpointの1つとして検討されている。QOLの評価には多くの方法が存在しているが、一定の信頼性と妥当性はまだ確立されていない。第Ⅲ相研究の主要評価項目はあくまでも延命効果であり、それを無視したQOLの議論は本末転倒であると思われる。JCOGの研究においても、医療者側が客観的に判断しうるパラメータでないと、経過を追ってみても正確なデータを集めることは困難と示されている。」

(西甲F32=東G58 p27)としており、この文献が出版された2004年10月時点においても、QOL評価における信頼性と妥当性は確立されていないことが示されている。

イ また、QOL評価にあたっては、「バイアスを防ぐためにはマスク化が非常に重要」である(西甲F30=東G56 p1216)。

この点、イレッサのドセタキセルとの比較国内第Ⅲ相試験では、イレッサが経口投与剤であるのに対し、ドセタキセルは点滴による静脈内投与であるため、マスク化、盲検化はできない比較試験であった(西福岡証人反対尋問調書=東丙G58 p21)。

このドセタキセルとの比較国内第Ⅲ相試験では、イレッサ群がドセタキセル群に対して症状改善を見たとされているが(西甲C5=東E10の3・18のスライド)、一方で「肺癌サブスケールによる評価」では両群で差はなかったとされている(同スライド)。前述のとおり、QOLの調査項目は、肺癌サブスケールによる肺癌の随伴症状の改善を含む形で、それ以外にも「友人を身近に感じる」というような抗ガン剤の治療効果以外の要素が極めて影響を及ぼすような質問項目で構成されており(西丙H44=東甲L157)、肺癌サブスケールにおいて改善を示さず、QOL評価でイレッサ群に改善が見られたとしても、少なくともマスク化、盲検化されていない同比較試験においては、QOLの改善はイレッサの治療効果でない可能性を否定できないと言わざるを得ない(西丙E49の1=東坪

井証人反対尋問調書 p 84～86)。

ウ さらに、「欠損値の処理も難しい。なぜなら、症状が悪くなるほど欠損値が発生する傾向にあるため、欠損値を無視して解析すると真の状態よりも見かけ上良くなってしまふという誤った結果につながることもあるからである。」(西甲 F 30 = 東 G 56 p 1216) とされる。

欠損、欠測とは、「進行肺癌などの予後不良な疾患においては、終末期に状態が悪くなり QOL 調査表に答えられなくなる。」(西丙 H 45 = 東丙 F 56 p 145) などのために、患者から回答を得られない状態を指す。

こうした欠損、欠測がある場合の解析については、種々の方法が提唱されているが、イレッサにおける ISEL 試験の QOL 評価においては、単に、測定が可能であった症例のみを取り出して評価されている(西福岡反対尋問調書 = 東丙 G 58 p 22, 23)。

しかし、こうした完全測定例のみでの評価は、「完全測定例における解析には、大きく2つの問題が生じる。1つは、欠測データが1つでもある患者を解析対象から除くため、解析に用いる対象患者数が大きく減少してしまうこと、そして、もう1つは、誤った結果を示してしまう危険が高くなることである。欠測データをもつ患者の割合が少なくとも5%未満でない限り、この解析アプローチを推奨できない。」(西甲 F 31 = 東 G 57 p 217) とされ、せいぜい5%の程度の欠測が許されるに過ぎない。そうでないと、患者背景等を調整してランダム割り付けした試験において、患者背景が調整されていない群間での比較がなされることとなってしまうからである。

この点、イレッサにおける ISEL 試験においては、イレッサ群 1129 例中 271 例が、プラセボ群 563 例中 138 例が欠測しており、明らかに5%を大きく上回る欠測があったにもかかわらず、なんらの調整もせずに評価されているに過ぎない(西福岡証人反対尋問調書 = 東丙 G 58 p

22, 23, ISEL試験の報告論文である西丙E34の8＝東丙G60の8においても、QOL評価にあたって何らかの統計的調整を行った旨の記載はなく、もし調整していれば当然記載されるはずであるところから見ても、ISEL試験のQOL評価は、単に完全測定例だけを取りだした評価に過ぎないことは明らかである）。

このようなQOL評価では、それ自体、信頼性に著しく欠けることは明らかである。

エ このように、QOL、症状改善等の患者の主観的なエンドポイントは、それ自体、主観的なバイアスを排除することが著しく困難なエンドポイントであると言わざるを得ない。

このため被告側証人である西條証人も、「第Ⅲ相研究の主要評価項目はあくまでも延命効果であり、それを無視したQOLの議論は本末転倒であると思われる。」（西甲F32＝東G58 p27）とし、また、「QOLの評価はあくまでも個人的な主観によってなされるため、第3者の価値観による分析は困難である。」（西甲F3＝東F15 p210）、「癌の治療効果判定にQOLの評価を持ち込むことについて、わが国では大きなボタンの掛け違いがあるように思えてならない。まず、QOLという言葉があまりにも安易に用いられていることである。QOLの定義、QOLの測定の方法論と妥当性、成績の評価と患者への還元、いずれをとってもあいまいなまま導入され、言葉が独り歩きしている点を指摘できる。更に危険なのは、QOLという言葉を利用し生物学的治療剤や経口抗癌剤の優位性を強調する動きが見られることである。経口抗癌剤を例にとれば、正確に服用した場合は効果もみられるが、毒性も注射によるものと同以上にみられている。また経口エトポシドのように静脈注射より毒性が強く効果が劣ることが経験されている。」（西甲F3＝東F15 p217）とまで言い切っているのである。

そして、JCOGではQOL調査はあくまで第Ⅲ相比較試験においてのみ調査され、検証的なプライマリーエンドポイントには用いず、探索的なセカンダリーエンドポイント（副次的エンドポイント）と位置づけられているのみである（西甲P125＝東甲L158・2枚目4d）。

平成17年改訂の抗ガン剤の新ガイドライン（西甲D5＝東甲H6）においても、同様にQOLは副次的な評価項目としての位置づけでしかない（西丙E49の1＝東坪井証人反対尋問調書p77）。

（5） 無増悪生存期間

なお、全生存期間をエンドポイントとした臨床試験における、後治療の影響を排除するためなどに、無増悪生存期間がエンドポイントとされることもある。

無増悪生存期間とは、「増悪が観察された」時点か、「あらゆる原因による死亡」の早いほうまでの時間として定義される（西甲F30＝東G56 p1214）。

この無増悪生存期間は、「生存期間のサロゲート・エンドポイントとして、無増悪生存期間が用いられることがある。」（西甲F30＝東G56 p1214）とされているとおり、生存期間という真のエンドポイントに対するサロゲート・エンドポイントに過ぎない。

このように無増悪生存期間は、腫瘍縮小の増悪をもエンドポイントとする点で、結局、腫瘍縮小効果とその継続期間を見ているに過ぎない結果ともなり得ることから、腫瘍縮小効果と同様の問題点を持ち、サロゲート・エンドポイントに過ぎないとされているのである。

しかし、無増悪生存期間をエンドポイントとする臨床試験においては、「マスク化していない試験では群間に偏りが生じる可能性がある。」、「すべての患者を定期的に評価することが必須である。つまり、増悪する可能性のあ

る部位を全て評価し、ベースラインと追跡の段階ですべての部位を完全に確認すること、それぞれの追跡時に同じ評価方法を用いること、同じ評価スケジュールを用いること、などが必要となる。これらの評価方法・間隔が群間で異なると実際には群間で無増悪生存期間に差がない場合でも見かけ上差が観察されてしまうことがある。」、「試験実施上の問題として、増悪が判断した時期だけを調べるのではなく、無増悪であることを確認した日付も記録しておく必要がある。これがないと増悪が確認されていない対象者に対する打ち切り時期を決定できない。」、「つまり、観察方法によって結果が異なる可能性があり、生存期間よりもソフトなエンドポイントといえる。」（西甲F30＝東G56 p1214）とされているように、増悪の評価等において、バイアスが生じる可能性が高いという問題点が指摘されているとおり、極めてソフトなエンドポイントに過ぎない点に注意しなければならない。

(6) NCI-PDQの記載

アメリカのNCI（米国国立癌研究所）では、PDQ（Physician Data Query）として、癌に関する情報を発信している（西甲F19＝東甲F35）。

その中で「癌治療研究に対する証拠レベル（成人）」として、エンドポイントの強さについても記述されており、「強さの降順に挙げる」として、まず、第1順位に「全死亡」、第2順位に「原因特異的死亡」、第3順位に「注意深く評価された生活の質」、そして、第4順位として「間接的な代替」つまりサロゲート・エンドポイントが挙げられ、その第4順位の中で、「無病生存」、「無増悪生存」、「腫瘍反応割合」の順に挙げられている。

そして、この間接的な代替については、「これらはすべて、研究者の解釈による影響を受けやすい。さらに重要なことに、それらは生存または生活の質などの直接的な患者の利益に自動的に変換されるものではない。」とされている。

このように、抗ガン剤の有効性評価の指標として最も重視されるべきは、全生存期間であることは、国際的なスタンダードなのである。

3 まとめ

以上のとおり、臨床試験のエンドポイントとしては、全生存期間—overall survival—が最も信頼に足る指標であることは明らかである。

このため、被告側証人である福岡証人も光富証人も、反対尋問において、全生存期間が最も重視されるべき指標であることを認めざるを得なかったのである（西福岡証人反対尋問調書＝東丙G 5 8 p 2 7，西光富証人反対尋問調書＝東乙L 2 4 p 3 7，3 8）。

もとより、福島証人、濱証人に加え、別府証人も全て、抗ガン剤の有効性は延命効果で見なければならぬことを証言しており、また、シルビオ・ガラティーニ教授もまた、全く同様に指摘している。同教授は、世界有数の研究所であるマリオ・ネグリ薬理学研究所所長であり、ヨーロッパ医薬品評価局委員や世界保健機構（WHO）の顧問なども歴任しており、レジオンドヌール勲章を初めとした国際的な賞も受けている高名な研究者である（西甲E 3 8＝東甲L 8 5 ガラティーニ意見書経歴欄，西甲E 3 9＝東別府証人主尋問調書 p 1 4，1 5）。

ガラティーニ教授もまた、抗ガン剤の有効性評価の指標として、以下のとおり指摘している。

「抗がん剤の有効性評価におけるプライマリー・エンドポイントは、もちろん、生存期間の延長である。既に述べたように、固形がんの種類は雑多な内容であるため、治療への反応か、，，本来の予後によるものかを予測することは難しい。このような条件のもとでは、反応率，無増悪期間，無増悪生存率といったエンドポイントは、せいぜい抗癌活性の評価指標とみなされる

に過ぎず、必ずしも臨床的有用性に関する正当な代用マーカールにはなり得ない。」（西甲E38＝東甲L85 p2）

こうしたガラティニ教授の意見を受けて、別府証人も以下のように供述している。

「先ほど申し上げたように、やはりプライマリーエンドポイントとして一番重視すべきは、やはり延命効果でございます、それ以外のものは非常に変わりやすいということでございます。ですから、例えばQOLを、その延命効果と並列的に並べて、こっちが駄目だけれども、こっちはいいというような、先ほど申し上げた二の矢三の矢としての使い方をするものではないというふうに思っております。」（西甲E39＝東別府証人主尋問調書p16）。

4 「抗悪性腫瘍薬の臨床評価方法に関するガイドライン」（「旧ガイドライン」西乙D7＝東乙H7）について

（1）はじめに

福岡証人は、「抗悪性腫瘍薬の臨床評価方法に関するガイドライン」の「第Ⅲ相試験成績においては、生存率、生存期間をプライマリーエンドポイントとし、他の適切なエンドポイントとして、QOLなどを求め、これらに対し、何らかの有用性（プライマリーエンドポイントが同等である場合は他の特徴を含めてよい）が示される必要がある。」（西乙D7＝東乙H7・11枚目）とされている部分について、主尋問において、「第Ⅲ相試験によって延命効果が証明されなかった場合、その抗ガン剤には有効性がないという趣旨か」という質問に対し、「何らかのほかの有用性が示されればそれでよいという理解ができる。」との証言をしており（西福岡証人主尋問調書＝東丙G57）、被告らも同様の主張をするようである。

しかしながら、このような解釈は、旧ガイドラインの第Ⅲ相試験の主要評価項目（プライマリーエンドポイント）の解釈を誤り、旧ガイドラインの要求した抗ガン剤の評価方法をねじ曲げるものでしかない。

（２） 旧ガイドライン注書（４）の記載

上記の旧ガイドラインの該当部分には、注書（４）が付されており、これによれば、「第Ⅲ相試験で目標としたプライマリーエンドポイントで同等性が証明された場合は、他の特性、例えばＱＯＬの改善（患者の肉体的苦痛の軽減、精神的満足度等）などの有用性が示される必要がある。」（西乙Ｄ７＝東乙Ｈ７・１５枚目）とされている。

すなわち、延命効果が統計学的に示されなかったとしても、ＱＯＬ等の何らかの有用性が示されれば良いという趣旨ではあり得ず、あくまで、延命効果で同等性が示された場合において、さらに他の有用性が示される必要があるという趣旨であることは明らかである。

なお、ここでの同等性とは、標準治療薬に対する同等性もしくは非劣性の証明を指しており、プラセボ、無治療・緩和療法群との同等性などではないことは、敢えて説明するまでもないであろう。比較対照群をプラセボ、無治療・緩和療法群においた時は、当然、被験薬は、それら比較対照群に対して優越している必要がある。

この点、光富証人は、あたかもプラセボ、無治療・緩和療法群に対する「同等」であっても良いかに述べるが（西光富証人反対尋問調書＝東乙Ｌ２４ｐ１３７）、偽薬であるプラセボや無治療・緩和療法群と生存期間が同等であるということは、被験薬に生存期間のベネフィットすなわち延命効果が無いということに他ならず、以下のような文献等の記載から見ても、旧ガイドラインがそのような趣旨で記載されていると到底考えられない。光富証人が本気でこのようなことを述べているとすれば、それは科学者として失格である

と言わざるを得ない。

(3) 解説文献等の記載

ア この旧ガイドライン作成班の班員である下山正徳氏は、旧ガイドラインの解説文献において、「延命効果をエンドポイントにして、安全性、有効性の他に、QOL(quality of life)などを加えて臨床的な有用性について評価する。しかし、QOLだけをエンドポイントにして、延命効果をエンドポイントにしないのは不可である。必ず延命効果があり、QOLもよいというのが最もよい。場合によっては、延命効果はそこそこ同等であるが、QOLが非常によくするというのであれば、これもよい。だから、必ず延命効果をプライマリー・エンドポイントにした試験計画書を出す。」(西甲D15＝東甲H10 p129)としている。

すなわち、作成班の班員である下山氏自身が、あくまで延命効果が存在することを前提として、他の指標を考慮するという趣旨を鮮明にしているのである。

イ 被告側証人である西條証人も、「癌治療における国際化」(西甲H13＝東甲F51)において、「Bridging studyの臨床的考察について」の項、右欄下から6行目末尾以下「薬剤の承認後、薬のsurvival benefitの確認のため、independentなphaseIII studyを二つ要求される。これでsurvival benefitがなければ承認取消となる。」(p211頁)と述べて、旧ガイドラインの承認後の第III相試験については、「survival benefit」つまり延命効果を確認するための「independentなphaseIII studyを二つ」つまり独立した2つの第3相試験が必要であり、これらで延命効果が確認されなければ承認取消となると明確に述べている。

同様に、西條証人は「癌の分子標的と腫瘍マーカーの開発」(西甲H14＝東甲F52)において、「第III相試験においては、over all survival (O

S) あるいはTTPを評価するため、これらに代わるサロゲート・マーカーは考えられず、**survival benefit**の有無を検討することがもっとも重要である。」

(p 87右3行目)、「臨床試験の早い段階でサロゲート・マーカーをもって有効性が示唆されても、最終的には第Ⅲ相試験で延命効果が証明されなければ臨床応用されないことはいうまでもない。」(p 88左8行目)と述べて、抗ガン剤の第Ⅲ相試験においては延命効果の確認が最も重要で、延命効果が確認されない限り「臨床応用」されないと述べている。

西條証人は、イレッサについても市販後第Ⅲ相試験において延命効果が示せなければ承認が取り消されるべきであることを認め(西乙E20＝東西條証人反対尋問調書p 113)、また、イレッサが統計学的には有用性を持っていないことを認めている(西乙E20＝東西條証人反対尋問調書p 130)。

(4) 別府証人の供述

別府証人も、旧ガイドラインにおける延命効果とQOLの改善等との関係について、以下のように供述している。

「例えば、延命が数箇月延びたといったします。しかし、それが非常に苦痛に満ちた内容であったとしたならば、それは非常にクオリティーオブライフ、生活の質としては非常に劣ることになります。で、そういうことがあってはならない、つまり数字の上だけの延命ではなくて、やはりその人にとって価値ある時間であってほしいということでありまして、QOLをここに加えた理由は、まさに、さらに詳しい、さらに厳しい条件をそこに加えているというふうに読み取るべきであると思います。いやしくもこれを、QOLというものを、例えば、延命効果はないけれども、QOLでこれだけ少しよくなったところがあったからというような、言ってみれば、

その効果がうまく当たらなかったときの、二の矢、三の矢、というような形で使うべきものではないだろうと思います。」（西甲E39＝東別府証人主尋問調書p10）。

（5） まとめ

このように、旧ガイドラインの趣旨は、あくまで延命効果が統計学的に証明されることが大前提となっていたものである。ただ、その証明が市販後に許されるという構造に過ぎない。

決して、延命効果が統計学的に証明されなくても、他のQOL等の指標によって有用性を判断しても良いなどということではあり得ないのである。

そのため、福岡証人も、反対尋問においては、この事実を認めざるを得なかった（福岡反対尋問調書p29）。

なお、福岡証人は、反対尋問に対して、「延命効果が同等でもいいですよね。」「延命効果が劣れば駄目ですよね。」等と述べている。これは、あるいは、イレッサにおけるドセタキセルとの比較国内第Ⅲ相試験において、「イレッサがドセタキセルに劣っていることが証明されたわけではない」という、独特の議論を前提にした回答であるとも考えられる。

しかしながら、同国内第Ⅲ相試験においては、イレッサのドセタキセルに対する非劣性というプライマリーエンドポイントが証明されなかったのであり、イレッサがドセタキセルに対して同等であるとか、劣っていないなどとは言えない状況であることは明らかである。そして、非劣性試験の性質上、イレッサがドセタキセルに「劣っている」、ないしはドセタキセルがイレッサに「優越している」などということは、元々証明対象、対立仮説ではなく、そのようなデザインはされていないのであるから、同試験の結果をもって、「イレッサがドセタキセルに対して劣っていると証明されたわけではない」

などという議論は、根本的に誤った極めて非科学的な議論に過ぎない。このような主張こそ、「有効性は確実に」「有効という証明がない限り無効と考
えなければ」という医薬品の有効性評価の大原則を踏みにじるものである。

被告らの主張は、別府証人が述べるように、延命効果に対する副次的評価項目に過ぎないQOL等の指標を、まさに「二の矢、三の矢」として利用しようとするものであり、このような考え方は、旧ガイドラインの趣旨に明らかに反し、到底受け入れられるものではない。

そして、これまで述べてきたように、抗ガン剤の有効性は基本的に延命効果で評価されるべきものである以上、旧ガイドラインにおいて腫瘍縮小効果の確認により承認が許される構造となっているととしても、延命効果の確認のないままの承認である以上、その際の有効性、安全性の評価は、より厳格になされる必要がある。この点については、さらに後述するとおりである。

第2 個別症例による有効性評価について

1 はじめに

被告らは、西被告会社準備書面（5）p 21以下＝東被告会社準備書面（5）p 20以下をはじめとして、イレッサの投与を受けた個々の患者についての症例報告の存在を列挙する。そして、被告側証人のうち、大阪地裁での西日本訴訟における光富徹哉証人、福岡正博証人、坪井正博証人もまた、いずれも、イレッサ投与例において著効例、スーパーレスポonder例があったなどとしている。

しかしながら、このような個別症例は、医薬品の有効性のエビデンス（証拠）たり得ず、これをことさらに強調することは、むしろ医薬品の有効性評価を見誤ることとなる。

以下、詳述する。

2 症例報告は医薬品の有効性のエビデンスたり得ないこと及びその理由

症例報告は、下記のようなバイアス、すなわち、医薬品の有効性が正しく評価されず歪められてしまうという現象を回避することができず、医薬品の有効性のエビデンスはない。

(1) 出版バイアス、発表バイアスを回避できない

出版バイアスないし発表バイアスとは、結果がうまくいったものについては出版、発表し、そうでないものには、出版、発表しない傾向にあることをいう。

この点、症例報告という手法は、報告者の主観により、結果がうまくいった症例を報告し、うまくいかなかった症例は報告しない、ということが容易に可能である。すなわち、症例報告は、「出版バイアス」ないし「発表バイアス」の影響から逃れることはできない（西光富徹哉証人反対尋問調書＝東乙L 2 4 p 5 6～5 7）。

(2) 選択バイアスを回避できない

また、ある症例や患者群を選択するにあたって、選択する集団や症例が母集団を正しく代表していないときに、そこで用いられる薬剤や治療法の評価を誤ってしまう事を「選択バイアス」という。例えば、予後因子のよい患者、すなわちもともと諸条件から治りのよい患者を選んでしまい、薬剤や治療法に関係なく疾患が改善した場合に、薬剤や治療法によって改善したのだと誤ってしまうことがありうる。

この点、症例報告という手法は、集団のなかから個別の症例をまさに選択して取りあげるのであって、「選択バイアス」から逃れることはできない（西光富徹哉証人反対尋問調書＝東乙L 2 4 p 5 7～5 8）。また、同一症例においても、そのうちの一部の情報のみ切り取り、例えば結果や背景因子等を選択的に情報として除外するという行われがちであるが、これもまた「選択バイアス」によ

るものである。

(3) 観察バイアスを回避できない

また、ある症例の観察者が、薬剤や治療法とその結果の関係について予断を有している場合に生じるバイアスのことを「観察バイアス」という。例えば、当該薬剤投与例に効果が出れば、それは当該薬剤ないし治療法によるものである、と考えがちである、という予断が考えられる。

これについて、元国立療養所院長の砂原茂一氏は、その著書において

「患者というものは、1人1人特別な条件をもっていて、2人として同じ患者というものは存在せず、1人の患者にある治療をほどこしたあとで病状が良くなるのが観察されたとしても、同じ病気の他の患者に同じクスリを与えたとき同じような効果が期待できるとはかぎらないこと、さらに病気には自然治癒というものがあって、くすりをつかわず、手術をしなくとも、自然によくなることがある」
「よくきくはずのくすりをつかったのだから、また、一生懸命治療したのだから、病気のよくなったのはそのくすりのおかげに違いないと、自分自身いいきかせているにすぎません。」（西甲G4＝東甲L49 p90～91）

と述べ、てるてる坊主や千人針の例を引いた後、

「つまり、時間の前後関係を因果関係ととりちがえるのです。こういう考え方の誤りはいうまでもありませんが、治療の場合には、しろうとのみでなく医者もしばしばこのような誤った判断に導かれます。」

と述べている。

すなわち、個別の症例を観察して、投与した医薬品に有効性があったと判断するのは、「飲んだ」「治った」「効いた」の「3た論法」として、医薬品の有効性を判断するにあたって、最も初歩的基本的な誤りなのである。

症例報告という手法は、観察者＝担当臨床医であり、当然、ある薬剤や治療法を使用したことを知っているのであり、「観察バイアス」に極めて影響を受け

やすい（西光富徹哉証人反対尋問調書＝東乙L24 p59）。

（４）症例報告のエビデンスレベル

NCI（米国国立癌研究所）－PDQのサイト（西甲F19＝東甲F35）には、ガン治療のエビデンスレベルについて、研究デザインの観点からその高低を順位づけた記載がある。これが、高い順から「１．ランダム化対照臨床試験」、「２．非ランダム化対照臨床試験」、「３．ケースシリーズ」となっており、３の中でも、「i 集団ベースの連続シリーズ、ii 連続したケース（集団ベースでない）、iii 非連続のケース」とさらに細かく高低が定められている。

この基準に照らして考えると、被告側証人が主尋問で行った症例報告は、それぞれ学会や論文等で別々に発表されたもので、連続したケースではないから、エビデンスレベルとしては最も低い3 iiiの「非連続のケース」である。なお、同基準には、「臨床経験は最も弱い形態の研究デザインである」と記載されている。

さらにいえば、さきほどのNCI－PDQの基準によれば、「研究または臨床経験はデザインの強さとエンドポイントの重要性の両方によって順位付けされる」とされている。被告側証人らは、単に症例報告をしたのみであって、集団ベースの腫瘍反応割合についてのデータを提供しているわけではないので、上記基準のD iiiにもあてはまらない。

これら、被告側証人らが本法廷で行った症例報告は、イレッサの医薬品の有効性のエビデンスとしては「なし」というのが端的な結論である。

（５）医薬品の有効性は臨床試験の結果によって評価すべきであること

このような症例報告等の観察研究が回避できない上記のようなバイアスを出来る限り除去し、医薬品の有効性を正しく評価するためになされるのが、比較臨床試験、とくにランダム化二重盲検比較臨床試験なのである（西甲F40「Evidenceと臨床試験」＝東甲F61）。

抗ガン剤の旧ガイドライン等をみれば明らかなおおり、我が国において、抗ガン剤が、臨床試験を行わず、症例報告やその集積を根拠として承認されることはあり得ない。これは、症例報告では、医薬品の有効性が正しく評価されず歪められてしまうというバイアスを回避することができないからにほかならない。

この点について、工藤翔二証人は、東日本訴訟における主尋問において、医薬品の有効性評価のあり方として、

「本当の全体としての有効性というのは、そういう患者さん全体をマスマで取り扱って、統計的に処理して生存期間がどれぐらい伸びるかとか、そういうことで判断していくものです。」（東工藤翔二証人主尋問調書＝西では未提出 p 1 1 2）

と証言する。これは、医薬品の承認の根拠となる有効性の評価は、統計的手法、つまり臨床試験によって行うべきであるとする趣旨である（東工藤翔二証人主尋問調書＝西では未提出 p 4 8～4 9, p 5 1～5 2）。すなわち、1つの症例報告の結果から、他の症例でどの程度類似の結果が得られるかというのは、症例報告そのものからは分からない。そうすると、どれだけ症例でどれだけ効果が出るのかというのを見るためには、全症例を見なければならず、結局、臨床試験によらざるを得ない（東工藤翔二証人主尋問調書＝西では未提出 p 4 9）。全く正当な指摘である。

また、個別症例が医薬品の有効性評価の根拠となり得ないということは、東日本訴訟の西條長宏証人の文献（東甲 F 5 9＝西甲 F 5 8『癌診療とEBM Part 2』巻頭言）においても、下記のとおり、強調されているところである。

「学会に出席して頻繁に目にする抄録のタイトルとして『当院における〇〇の治療成績』あるいは『当院における〇〇診療の現状』がある。これらのお大半は治療計画書（プロトコール）なしに、当然IRB（Institutional Review Board）の審査もなくIC（Informed Consent）formもなく、clinical practiceとして行われたきわめて多岐にわたる診療（治療）による奏効

率や、生存期間をretrospective にまとめたものが多い。これらは当然エビデンスレベルとしてはきわめて低いものであり、学問の進歩に寄与するものではない。また、これらの発表の最後に演者は必ず『今後さらに症例を重ねて検討したい』としめくくることが多いが、症例を重ねても何ら得ることがないことは自明である。」

これもまた、正当な指摘である。

以下、これに対して、医薬品の有効性の根拠として本法廷で症例報告を紹介している光富徹哉証人、福岡正博証人、及び坪井正博証人の個別の証言内容を確認する。

3 光富徹哉証人について

(1) 光富証人の証言内容及び目的

光富証人は、その証言の相当部分を自ら医師として経験した症例報告の説明に費やしている。光富証人が、別症例報告をもって医薬品の有効性の根拠としようとする意図のもとに証言をしていることは、光富証人自らが「示唆に富む症例を紹介することにより、医療現場にたずさわらない方々にゲフィチニブによる肺がん薬物療法の実地臨床における実態と特徴を理解してもらうことに加え、ゲフィチニブに関する一専門家の医学的評価を提供するためである」（西乙E 1 2 = 東乙L 1 4 p 1）と述べているとおりである。すなわち、光富証人は、イレッサがよく効く（と思われる）個別症例を選択して報告し、イレッサの「医学的評価」の根拠としようとしているのである。

(2) 光富証人の症例報告もやはりバイアスを回避し得ない

しかしながら、光富証人の症例報告も、やはり、上記述べたような各種のバイアスを回避し得ていない。

光富証人は、「症例報告をするに当たっては、診療行為中に見られた事象

のすべてを取り上げる必要はなく，自ら提供しようとする医学的評価を形成する上で必要かつ十分と考える範囲で，その一部を取り出し，整理することになる。」（西乙E 1 2＝東乙L 1 4 p 2）と，「選択バイアス」があることについて自ら認めている。法廷において紹介した6症例のうち，症例1・4以外の4例について死亡したことすらも当初は触れず，投与後いつ亡くなったかなどは意図的に情報として除外している。また，他方で，光富証人は，イレッサの毒性については，ほとんど全くといっていいほど触れていない。わずかに，意見書において「一方で，副作用として間質性肺炎を発症した患者は，当診療科では100症例強中3症例を経験しており，そのうちの2症例は間質性肺炎によりお亡くなりになられた。」とあるのみで，その症例経過も，患者背景因子も全く分からないし，証言にあたってこれを確認検討したという形跡すら見受けられない。

光富証人自身も，自己の文献において，「専門誌に掲載される論文には，ポジティブデータが選択される傾向となる（出版バイアス）。また，研究者としては，ポジティブデータを発表したいために，ある場合は特殊な条件下のポジティブデータを売りとして発表しがちとなる」（西甲H 1 5＝東甲F 6 2）と述べているところであるが，上記のとおり，光富証人は，本法廷における証言においても，自ら「示唆に富む症例」をセレクトした，としており，出版バイアスの影響にあることを自ら認めている。光富証人自身も，「エビデンスレベルからいえば臨床試験にそれに比べて低いものであることは言われるまでもなく十分承知している」（西乙E 1 2＝東乙L 1 4 p 5）と述べているところである。

4 福岡正博証人について

福岡正博証人は，「第I相試験の対象になるような非常に進行した非小細胞肺癌患者について，イレッサが著効を示したことは，正に驚くべき経験であった」（西丙E 3 3＝西丙G 5 9 p 9）と述べ，また，主尋問においても，スーパーレス

ポonderという言い方をして、イレッサの有効性の根拠の一つとして、こうした患者さんが存在していることを挙げている（西福岡正博証人主尋問調書＝東丙G 5 7 p 3 5）。

しかしながら、このスーパーレスポonderという趣旨不明のことばを用いて、個々の著効例が医薬品の有効性の根拠となるという考え方自体、医薬品の評価をゆがめるものである。すなわち、既に上記2（3）で述べたとおり、個々の症例のみをみて、スーパーレスポonderがある、これをもってイレッサが効いていると判断することは、医師の観察バイアスに歪められた判断でしかないのである。

なお、未承認薬であるが、癌患者について著効例の症例報告がいまも後を絶たないものとしては、丸山ワクチンが有名である（西甲P 7 9＝甲L 1 4 3）。福岡証人は、丸山ワクチンについての著効例の存在をもってその有効性を評価することができないとした上で、イレッサの著効例なるものについては、「丸山ワクチンと一緒にしてもらっては困る」などと述べながらも、自らもまたイレッサについて全く同じことを述べていることにはたと気づいたのか、イレッサの著効例は、イレッサにとって有効性の根拠とはならず、「一つの情報」（西福岡正博証人反対尋問調書＝東丙G 5 7 p 4 7～4 8）に過ぎず、検証されていない、ということ認めざるを得なかった。

このように、福岡証人のいうスーパーレスポonderなるものは、イレッサの有効性の根拠となり得ない。そればかりか、仰々しい言葉でもって、イレッサの医薬品評価を見誤らせる有害な「情報」として悪用され得るものといわざるを得ない。

5 坪井正博証人について

そして、坪井正博証人も、また、東日本訴訟における証人尋問において、実臨床での症例を取り上げて、イレッサの有効性評価を行おうとしている。

例えば、坪井証人は、イレッサについてのV-15-32試験の中の1症例を取り上げ、「延命効果があるであろうと強く推察しています」とする。

しかしながら、坪井証人は、反対尋問においては、主尋問で取り上げた経験症例3例は、いずれも学会や論文などで公表したものですらなく（西丙49の1＝東坪井正博証人反対尋問調書p55）、臨床医の目から見て個々の患者にとってベネフィットがあったと思われる症例を紹介したものにすぎず、この3症例をもって、「医薬品の承認の根拠としての有効性」の根拠とはならないことを自認している（西丙49の1＝東坪井正博証人反対尋問調書p56）。

また、同時に、坪井証人は、V-15-32試験において取りあげた1症例では確かに効いているようにみえても、全体の結果として、延命効果（非劣性）は証明されていないということは、他方では短期で死亡されてしまう症例があるためであると考えられるからであることを認め（西丙49の1＝東坪井正博証人反対尋問調書p57～58）ている。

このように、結局のところ、坪井証人は、医薬品の有効性評価にあたっては、臨床試験の対象患者総体において延命効果があるかどうか重要なのであり、その一部の個別症例を取り上げてイレッサの有効性評価を行うのは適切ではなかったことを認めざるを得なかった。

他方で、坪井証人は、V-15-32試験での間質性肺炎発症例3例の経過については何ら答えなかった（西丙49の1＝東坪井正博証人反対尋問調書p58）だけでなく、イレッサ投与群の中から、短期間で死亡された症例や、重篤な副作用が発生した症例を取り上げて、「イレッサは延命効果どころか余命短縮効果がある」と結論づけることは妥当ではない。それは仮説である旨述べた上で、原告代理人の「1例だけ取り上げて、治療効果があるというのも仮説なんじゃないですか。」という質問には事実上答えず、ごまかす（西丙49の1＝東坪井正博証人反対尋問調書p59）など、イレッサについては、なりふりかまわず、公平性を欠く証言態度で本件訴訟に臨んでいることが明らかであ

る。

6 まとめ

以上述べたとおり，個別症例の存在を医薬品の有効性の根拠としようとする被告会社の主張，及び光富証人，福岡証人，坪井証人の各証言は，いずれも症例報告である以上バイアスを免れ得ないものであり，失敗に終わっている。エビデンスのないものをいくら集積してもエビデンスにはならない。医薬品の有効性は，上記第1で述べたとおり，比較臨床試験の結果によってはじめて証明されるものであり，臨床試験結果が出ているにもかかわらず，個別症例をもって医薬品の有効性を論じようとする事自体，医薬品の適正な有効性評価を妨げるものでしかない。

なお，これら被告側証人のなした症例報告は，いずれもイレッサ承認後のものであり，イレッサの承認時までに審査資料とされていたものではないことについても念のため付言しておく。

第3 奏効率による延命効果の予測の問題性

1 序論

(1) 被告らは，承認時におけるイレッサの有効性の最大の根拠として，IDEAL-1の結果，特に日本人群における奏効率（正確には「反応率」）の存在を挙げている。

しかし，既に述べたように，奏効率は，延命効果のような真の臨床的ベネフィットを計測するものではなく，延命効果の代替エンドポイントに過ぎない。更に言えば，一定程度の奏効率が確認されたとしても，これをもって薬剤の延命効果が直ちに推測できるというものでは全くなく，有効性の見込み

判断においては極めて慎重な態度が求められる。

被告側申請の西條証人は、自身が執筆者として名を連ねる論文において、「奏効率を根拠に加速承認を行うことが妥当であるかどうかについて検証されていない」と指摘し、「医薬品行政には、…リスクやハームを上回るだけのベネフィットがない新薬が出回ることを防ぐことが求められているわけだが、現在の日本の制度がそれにも合ったものであるとは言い難い。」として、批判的な見解を述べている（西甲H18＝東甲G55 p586）。ここに述べられるような、奏効率を根拠とした第Ⅱ相承認の問題性については、論文執筆時（2004年）のみならず、イレッサ承認当時においても同様であり、奏効率のみに依拠して安易に有効性の見込みを肯定するようなことは許されない状況にあったものである。

(2) 本項においては、以下の観点から、奏効率による抗がん剤評価の問題性について整理し、奏効率の高さをもって直ちに延命効果が予測されることが誤りであることを明らかにする。

- ① まず、高い奏効率を示しても延命効果に繋がらないという問題性はイレッサの承認、販売開始以前に明らかとなっており、このことは承認時のイレッサの評価にあたって大前提とされなければならないということである（後述「2」参照）。
- ② 次に、このような問題性は、「奏効率」の判定方法が第Ⅱ相試験段階におけるスクリーニングの目的に沿うように定義されていたことによる当然の帰結であるということである。この点からも、抗がん剤の第Ⅱ相試験結果について、第Ⅲ相試験へと進むスクリーニングの意義を超え、市販承認の前提として延命効果を予測しようとする際には慎重さが求められる（後述「3」参照）。
- ③ この点について、被告側申請の証人らは、奏効率と生存期間中央値に統

計学的な相関関係があることをもって、延命予測の論拠とする。しかし、そもそも、このような相関すらなければ奏効率も代替エンドポイントたり得ないのであって、相関があるからと言って、奏効率が延命効果を十分に予測させるかどうかを合理的に明らかにしたことはない（後述「4」参照）。

④ 前記③について更に敷衍すれば、被告らが根拠として挙げる福岡論文（西丙E34の5＝東丙G60の5）、西條論文（西乙H38の2＝東乙F11の2）を見てみると、そもそも、相関関係の存在についてすら、極めて不十分な確認にとどまっているものであるから（後述「5」参照）、奏効率と生存期間中央値との相関に関する被告らの主張は、それ自体が根拠を欠くものである。

⑤ そして、奏効率が精度の低い延命予測しか提供できないことは、抗がん剤承認の全体の制度の枠組みにおいても、当然の前提とされている（後述「6」参照）。

2 奏効率が延命につながらない事例がイレッサ承認前に報告されていたこと

(1) そもそも、第Ⅱ相試験において一定の奏効率が観測されていたのに、第Ⅲ相試験で延命効果の検証に失敗した例は枚挙にいとまがない。奏効率が延命につながらない点については、既にイレッサ承認以前の知見となっていた。

(2) この点に関しては、被告側証人である西條医師自ら、「従来の抗がん剤で、第Ⅱ相試験での奏効率の大小では第Ⅲ相試験での生存のベネフィットを予測できないことが指摘されてい」たことを認める（西乙H38の2＝東乙F11の2 p582（西乙E20＝西條証人反対尋問調書p65で確認））。福岡証人も、多くの臨床試験の報告では、奏効率は新抗がん剤群で高かった

が、生存がないという結果が存しており、非小細胞肺癌に対する治療評価は奏効率では不適切で、生存率、QOLのような真のエンドポイントの改善が重要と断じている。（西甲H10＝東甲F56（西福岡証人反対尋問調書＝東丙G58 p36で確認））。この文献においては、表4（p86）、表5（p87）において、この当時報告されていた試験結果について、一般的に紹介されている。ここに挙げられた合計12の比較試験のうち、実に5つの試験において、より高い奏効率が見られた群において、生存期間中央値（被験群の半数が死亡するまでの期間を指す。「MST」との略称も用いられる。）が短くなるという「ねじれ現象」が見られている。下の表は、「ねじれ現象」を示した5試験を抜粋したものである。このうち、Giacconeの試験においては、18%もの差で高い奏効率を示した群において、延命とは正反対の結果となる、生存期間中央値の短縮が見られた点で注目される。

表	報告者	レジメン	症例数	奏効率	奏効率の差	生存期間中央値
4	Gatzemeier	Cisplatin/Cisplatin+Paclitaxel	206/202	17/26	9 %	8.6/8.1m
5	Giaccone	CDDP+PCT/CDDP+VM16	166/166	47/29	18 %	9.4/9.7m
5	Belani	CBDCA+PCT/CDDP+VP16	190/179	22/14	8 %	7.7/8.2m
5	Crino	CDDP+GEM/CDDP+MMC+IFO	154/152	40/28	12 %	35/38wk
5	Niho	CDDP+CPT/CDDP+VDS	100/103	29/22	7 %	45.4/49.6wk

この外にも、2000年の小細胞肺癌の研究において、既に、「第Ⅱ相試験で有望と思われた試験レジメンが、生存期間につながることはまれにしかないことが明らかとなった」ことが指摘されていた（西甲H27＝東甲F53、チェン論文（西乙E20＝西條証人反対尋問調書 p65で確認））。

また、西條証人も、この文献を引用して、「奏効率が生存のベネフィットを示唆する適切な代理エンドポイントであるかという点には議論の余地がある。」と述べているのである（西甲H18＝東甲G55 p582左段）。

なお、西條証人は、このチェン論文は小細胞肺癌を対照とするもので、非小細胞肺癌には当てはまらないかのように証言するが（西乙E20＝西條証人反対尋問調書p65以下）、自身の論文においてそのような限定を施していないことは明らかで（西甲H18＝東甲G55）、かかる反論は、自らの立場を場当たりの変遷させた供述に過ぎない。

(3) このように、既に報告されていた実例から、奏効率が抗がん剤の有効性である延命効果を予測させる代替エンドポイントとして必ずしも妥当ではないということは、イレッサ承認以前において既に明らかになっていたことであった。

3 奏効率の判定基準による問題性

(1) 奏効率の判定基準に内在する問題

一定の奏効率が確認されても、それが延命につながらないという問題の理由としては、予後因子バイアスやタイムバイアスといったバイアスの存在も指摘されるが、この点は後述することとして、ここでは、まずもって、奏効率の判定基準そのものに内在する問題があることを指摘する。

I D E A L 試験で採用された腫瘍縮小効果判定基準は、修正WHO基準と呼ばれるものであって、部分反応（PR）に該当するには、計測腫瘍の50%以上縮小が4週間以上継続することとされていた（西丙C1＝東丙D1申請概要p460、西乙E18＝東乙L10西條意見書p10～11）。なお、現在はRECIST基準（西甲G7＝東甲L47）が広く用いられているところ、その実質においてWHO基準と大きく変わるものではない。

しかし、この基準は、延命効果を予測するための代替エンドポイントの基準としては緩やかに過ぎるのであって、全く十分なものではないのである。

この点は、被告側申請の西條証人自身が、自らの論文で、「腫瘍の大きさで50%以上減少することは腫瘍細胞の最小殺傷のみを意味しており、全身腫瘍組織量において生物学的に意味のある減少を必ずしも示唆するものではないことである。加えて、相対的に短い反応期間（1ヶ月以上）は、生存期間を有意に延長するには不十分であると思われる。」と指摘している（西乙H38＝東乙F11訳文5）。

（2） 延命の予測精度の低い基準が採用された理由

抗がん剤の第Ⅱ相試験において、このような緩やかな基準が採用されたのは、第Ⅱ相試験は、単に、効果のない薬剤を早期にスクリーニングするという「探索」目的の試験として位置づけられるからである（西原告第15準備書面p14以下、西甲F46＝東甲F80訳p2右段参照）。

この点、通常の承認過程においては、第Ⅱ相の後、市販承認される前の段階において、第Ⅲ相における延命効果の「検証」が行われることが前提となる。このため、第Ⅱ相段階では、第Ⅲ相に進むか否かのスクリーニングができれば十分である。この予備的な段階で延命予測の精度を上げることにのみならず、捕らわれて過大な労力や時間を割いてしまつては、制度全体として非効率・不経済となってしまう。そこで、第Ⅱ相で用いられる「奏効率」という指標についても、延命予測の精度をある程度犠牲にしても、より効率よく第Ⅲ相試験へのスクリーニングを行えるように設定すべきとされ、専門家の議論を踏まえて、上述のような基準が策定されたものである。

このように、現在、一般に用いられている「奏効率」という指標は、延命効果は後の第Ⅲ相で確認するという大前提のもと、延命予測の精度を一定犠牲にしながら設定されたエンドポイントということができる。

(3) 小括

以上のような「奏効率」の性質からして、第Ⅱ相試験の奏効率の結果のみをもって真のベネフィットの証明を代替できたと誤信し、安易に有効性を肯定することは許されない。腫瘍縮小効果の判定に用いる基準そのものに内在する限界がある以上、ある試験で一定の奏効率が認められたとしても、その結果から延命効果を予測するには慎重な検討が必要なのである。

4 相関のみでは、代替エンドポイントの「妥当性」確認には不十分であること

(1) 序論

ところが、本訴訟において、被告らは、奏効率と生存期間中央値に単純な相関があることをもって、奏効率が、延命効果を予測するのに有益な代替エンドポイントであることの論拠とし、被告側申請にかかる西條・福岡各証人もまたそれに沿う証言をしている。

しかし、かかる主張及び証言は、以下の点において完全な誤りである。

① まず、被告らが言う統計学的な相関の点をもって、奏効率が延命効果の代替エンドポイントして成立することには全くなならない。そのような統計学的相関すら認められなければ、もはや代替エンドポイントたりえないという意味では必要条件ではあっても、相関さえあれば十分に代替エンドポイントたりうるという十分条件ではないのである。被告らの主張は、あえてこの点を混同している。

② また、被告側証人らが指摘する統計学的相関の根拠とする論文には、相関の確認の限度においてさえも問題があり、そもそも、かかる相関があることの根拠自体に問題があるといえる。

本項においては、上記①の点について述べ、上記②については項を変えて後記5項において整理することとする。

(2) 相関のみでは代替指標としての「妥当性」は認められないこと

ア 代替指標としての「妥当性」(バリデーション)について

一般に代替エンドポイントの候補となる指標について、真のベネフィット予測の精度の存否を客観的に確認する統計的手法は、この分野では一般に「バリデーション」, 「妥当性研究」などと呼ばれている。先に述べたように、真のエンドポイントを推測させる代替指標として成立するには、最低限の必要条件を満たしていることは当然として、更に十分条件も満たさねばならない。この点については、山本精一郎氏も「サロゲートエンドポイントを用いるためには、それが上に挙げたような統計的な要件、臨床的な要件を満たしていることを証明してから用いる必要がある。これを妥当性研究と呼ぶ。」とする(西甲F 30 = 東甲G 56 p 1212右段)。本書面においては、上記した点から、かかる両条件を合わせて、代替指標としての「妥当性」と表記して、以下論ずる。

代替指標の「妥当性」の確認に関しては、いくつかの手法が提案されている。しかし、そのいずれにおいても相関関係の確認のみで足りる、とするものはない。例えば、代替指標の「妥当性」確認の基準として、最もよく知られて繰り返し引用されるものがPrentice (プレнтаイス)氏が提案するものである。旧ガイドラインにも引用される生物統計学者Fleming (フレミング)は、この基準について、

「…(1) 生物マーカーが、臨床エンドポイントと相関関係を有すること、(2) そのマーカーが、介入の臨床効能エンドポイントに対する最終的効果を十分に捉えられることである。条件(1)だけで代替エンドポイントの正当性を実証するのに十分だと勘違いしている者も多いが、条件(2)のほうが充足しにくく、ずっと実証が困難なのである。(西甲F 47 = 東甲G 98 訳文p 4。なお、この訳文におい

ては、原文中の「validation（バリデーション）」について「正当性の裏付け」という訳語を当てている。）。」

と整理し、別の文献においても、「代替エンドポイントの検証にあたって、この相関関係は、必要条件ではあるが、十分条件ではない。」と繰り返し強調している（西甲H60＝東甲G106 訳文 p 1 右段）。

この点に関しては、RECIST基準の筆頭執筆者であるTherasse（テラッセ）氏による文献（西甲F46＝東甲F80 訳文 p 5 右段）を初めとして、各種学術文献においても繰り返し指摘が行われている。また、外国文献のみならず、上述の山本精一郎氏による日本文献でも、「単に相関があるだけでは、この（挿入： サロゲートエンドポイントを用いる必要）条件を満たさない。」と明記されている（西甲F30＝東甲G56 p 1 2 1 2 右段）。

このように、相関関係は、「妥当性」の必要条件とはなるものの、相関のみでは十分条件たりえない点については、明らかに、各種専門文献に一致する一般的な知見となっていたものである。

イ 西條、福岡証人らに共通して見られる混同

他方、西條証人も、意見書において「実証（validate）」という、同じ言葉を用いているものの（西乙E18＝東乙L10 p 10）、その引用する文献（西乙H38の2＝東乙F11の2）の内容を見ると、奏効率と生存期間中央値との間の単なる相関の検討以上のものを含むものではない。これは、代替エンドポイントの「妥当性」の確認という文脈においては、完全な誤用と評価しうるものである。（福岡証人も、相関のみを検討する西丙E34の5＝東丙G60の5に依拠して、奏効率からの延命予測を行っている点で、同様の誤りを犯しているといえることができる。）。

なお、西條証人が意見書において、

「なお、近年は、…腫瘍縮小効果と延命効果とが必ずしも相関すると

は限らないという考え方も見られる。…イレッサの承認当時においては、がんの専門医のコンセンサスは…腫瘍縮小効果と延命効果は相関するというものであり、…承認当時のかかるコンセンサスが合理的な根拠に基づいていたことを否定するのは、明らかに適切ではない。

（西乙E18＝東乙L10p13）」

と述べている箇所に関しては、反対尋問において、西條証人が根拠として挙げる自らの論文で確認したのは、奏効率と延命効果の相関などではなく、単に、奏効率と生存期間中央値の相関に過ぎないことを認めた（西乙E20＝西條証人反対尋問調書p64）。この証言からも、西條証人において概念の整理が不十分なまま混同していることが垣間見える。

(3) 相関の確認のみでは不十分である理由～タイムバイアス、予後因子バイアスについて

ア はじめに

このように、奏効率と生存期間中央値の間に相関関係が存在したとしても、奏効率が必ずしも延命効果の代替とはならない、即ち、相関関係の存在のみで代替指標としての「妥当性」を有することにならないということは、既にイレッサ承認以前に確立した知見というべきであった。

その実質的な理由として、奏効率判定における予後因子バイアス、タイムバイアスというバイアスの影響が挙げられる。この点は、西條・福岡両証人も認めるとおりである（西乙E20＝西條証人反対尋問調書p62，西福岡証人反対尋問調書＝東丙G58p34）。

イ 予後因子バイアス

ここで、予後因子バイアスとは、一般には、各患者の持つ予後因子（全身状態（PS）、喫煙歴、治療組入期間等）を予め均質に各被験群に振り

分けること（無作為化）をしないで、群を単純比較すると、各群の予後因子の差異の影響により見かけ上の違いがもたらされてしまうことを指す。そして、奏効群と非奏効群を比較する場合などには、予後因子と密接に関連する結果から群の振り分けを行うことにより、各群の予後因子に偏りが生じてしまい、見かけ上の薬剤の効果があるかのように観察されてしまう。さらに、複数の試験結果を対象とした相関の確認についても、上にも引用したBuyseの文献にも説明されるように、

「ある研究で、ほとんどの被験者の疾病が非常に進行して全身状態が不良であれば、低い奏効率と短い生存が観察されると予想される。反対に、ほとんどの被験者の限局期にあり全身状態が良好であった試験では、高い奏効率と長い生存が観察されると予想できる…（西甲H 6 1 = 東甲F 9 0 p3 ~4）」

ため、生存と奏効の双方に作用する予後因子の影響により、薬剤の効果の有無に関係なく生存と奏効の相関関係が得られてしまうことになるが、これも予後因子バイアスの影響として整理することができる。

この点に関して、西條証人も、反対尋問において、ある特定の被験者群における生存期間中央値は、薬剤の活性とは無関係に、状態のいい患者が集まれば長くなり、状態の悪い患者が集まれば短くなる可能性を認める。そして、予後的に良好な患者のみのデータ蓄積により、その薬剤活性とは無関係に長期の生存期間中央値となる場合がある、と明言している。（西乙H 3 8 の 2 = 東乙F 1 1 の 2（西乙E 2 0 = 西條証人反対尋問調書 p 6 3 において確認））。

ウ タイムバイアス

次に、タイムバイアスとは、「奏効者に関しては、“奏効（response）”が観察されるまで生存していることが必要なので、治療による延命効果の

あるなしに関係なく、非奏効者よりも生存時間は長くなるというバイアス」のことをさす（西甲F 3 3=東甲F 5 4 p 1 6（西福岡証人反対尋問調書=東丙G 5 8 p 3 4）。）。この場合も、予後因子バイアスと同様、薬剤の効果に関係なく生存と奏効の相関関係が得られてしまう（なお、福岡証人は、多数の試験を対象とした解析についてはあたかもバイアスの影響を受けないかのように証言しているが（西福岡証人反対尋問調書=東丙G 5 8 p 3 5以下）、その根拠が明示されることなく、西條証人の見解とも相違している点で、信用性は皆無である。）。

当然のことながら、このような相関をどれだけ精密に確認したところで、薬剤の効果から独立したものである可能性がある以上、奏効率の代替エンドポイントとしての「妥当性」の根拠にはならない。すなわち、ある薬剤投与群において一定の奏効率が得られても、論理的に、延命効果が予測されることにはならないのである。

（４） Buyse論文による相関分析の批判

ア 相関のみに依拠して代替エンドポイントを評価する手法の問題点

生物統計学者Buyseの論文（西甲H 6 1 =東甲F 9 0）は、西條氏の分析（西乙H 3 8 の 2 =東乙F 1 1 の 2）、福岡氏の分析（西丙E 3 4 の 5 =東丙G 6 0 の 5）のような、奏効率と生存期間中央値の相関についての分析を正面から批判する内容となっている。

ここでは、

「複数の試験がある場合、腫瘍縮小と生存の関係の評価方法として直観的に訴えやすい（が、重大な誤解を招きやすい）方法は、各研究の奏効率を、（例えば生存期間中央値といった）生存の要約データと対比させてプロットし、これらのポイントを通じて回帰線に適合させるやり方である。…回帰分析は、これらの疾病の中で行なわれた臨床試

験で報告された奏効率と生存期間中央値に基づいて行われた。有意な回帰スロープは、生存の延長は奏効率の向上と相関していることを意味すると解釈された。」

と福岡・西條各証人が行ったものと同種の解析の存在を概観したうえで、
「このアプローチの背後にはいくつかの重大な統計学的問題がある。」

と断じている（西甲H61＝東甲F90 訳文 p 2）。その理由として、この後の記述において、上述した予後因子による問題の外、群としてのデータを用いた解析において、当該群を構成する個別患者における個体差が捨象されてしまう問題、といった点を指摘する。

イ Buyse論文の位置づけと、証人らの証言の信用性が低いこと

このBuyse論文（西甲H61＝東甲F90）は、実は、西條証人の上記文献（西乙H38の1＝東乙F11の1）を、同一の掲載誌上において評釈するGelmon氏の論説（西乙H38の2のp1の目次の左上参照）「主要評価項目の明確なポイント： 肺がんにおける第Ⅱ相臨床試験」において、
「BuyseおよびPiedboisは、ある極めて重要な論文において、…このような調査の限界を注意深く概説した」ものとして引用されているものである（西甲H62＝東甲F91 訳文 p 2 左段）。したがって、この分野の研究としては良く知られた内容であり、専門家証人である以上、西條・福岡両証人も、一般的な知見として、当然知っていなくてはならないものである。特に、Gelmon氏に自らの論述を論評されている西條証人においては、この文献を確認すべきは当然である。

にも関わらず、本件訴訟における両証人は、Buyse論文が問題提起する上記の観点を全く無視するものであり、その信用性は低いと言わざるを得ない。

(5) 小括

以上より，福岡，西條両証人が，奏効率による延命予測の根拠として，相関の存在のみを挙げている点については，この主張の当否の検討を待つまでもなく，主張自体において失当で，不完全なものであることに留意されるべきである。

5 被告らが根拠とする研究からは相関の実証すら不十分であること

以上より，西條・福岡各証人は，奏効率による延命予測の根拠として相関のみを挙げる点において，主張自体失当との評価を受けても仕方のないものである。さらに，両証人が挙げる2つの文献の内容を精査してみると（西條証人の東甲F 1 1の2＝西乙H 3 8の2，福岡証人の西丙E 3 4の5＝東丙G 6 0の5），単なる相関の確認の限度においてすら，実証ありとは到底言えないようなものであった。上述のとおり，相関は，「妥当性」を論ずるにあたって必要条件として当然にその確認が必要となるものであるが，被告らが論拠とする統計解析については，この限度においてさえも十分な根拠がない，ということになる。以下，詳説する。

(1) 福岡論文・西丙E 3 4の5＝東丙G 6 0の5について

まず，福岡証人が執筆者の一人となっている丙E 3 4の5＝東丙G 6 0の5について検討する。

この文献における統計解析の対象とされた各数値は，各文献中に報告された数値（要約データ）をそのまま用いたものである。各試験，各患者について，背景因子の調整，症例登録数による重み付けはされていない。例えば，表1，2（丙E 3 4の5の2，p 8～9）の症例数を見ても，

・わずか17症例の群（表1・下から17行目），

- ・ 288 症例の群（表 1・下から 5 行目），
- ・ 488 症例もの群（表 2・下から 4 行目）が，

それぞれ，対等な 1 つの群としてしか評価されていない点だけを見ても，果たしてまともな結論が導かれるものなのか，大きな疑問を抱かずにはいられない。

さらに，この解析については，

「今回の調査には，多くの不均質な臨床第 II 相試験が含まれている。これらの試験では，症例登録数に大きな差があり，また患者背景の多くは不均質で，試験前に受けた治療法の数にも大きな差がある。奏効の評価方法にも差がある。それによって誤った結論が導かれる可能性がある。」

と，執筆者自身認めているものである（西丙 E 3 4 の 5 = 東丙 G 6 0 の 5 p 2 8）。

このような精度の低いデータを対象として，相関に関する統計解析を行ったところで，到底，相関関係の存否を検証したなどとはいえないことは明らかである（西福岡証人反対尋問調書 = 東丙 G 5 8 p 3 2 参照）。

（2） 西條論文（西乙 H 3 8 = 東乙 F 1 1）について

次に，西條証人自身も執筆者であり，相関の論拠とするこの解析において，まず第 1 に着目すべきは，無作為化第 II 相臨床試験では奏効率（訳文における「RR 値」）が生存期間中央値値（生存期間中央値の値）と相関していないという事実である。これらの治験では，生存期間中央値は，奏効率（RR 値）とは無関係に一定していることが示されている（西乙 H 3 8 の 2 = 東乙 F 1 1 の 2 p 4 左段）。この論文において，わざわざ，無作為化第 II 相試験を，他の単アーム試験と区別して解析したのは，無作為化試験は各群間の各種背景因子の調整なども要することから，結果の信頼性がより高いと考えら

れたためである（p 5 左段参照）。にもかかわらず、相関が示されなかったということは、仮に他の信頼性の低い試験の解析において相関が示されていたとしても、これが疑わしいということを示唆するものである。

第2に、本件論文中に「ある特定のRR値（奏効率）で広範囲の生存期間中央値を示すことは今後の課題を含んでいる。」（p 5 左段）と明記されていることからわかるとおり、この解析において確認された相関の精度は極めて低いものである。これでは、ある一つの群でRR値の増加が見られたとしても、生存期間中央値の延長を合理的に予測することなどできない。例えば、この分布図（p 4・図1）を見ればわかるとおり、奏効率が10%の群と、20%の群があったとしても、10%の群の生存期間中央値が9ヶ月ほどあり、20%の群の生存期間中央値が6ヶ月ほどしかなく、実際には奏効率の増加が余命の短縮をもたらすといった極端な事態が生じる可能性ですらも排斥できない。

当然のことであるが、「妥当性」の必要条件としての相関関係の確認にあたっては、単に弱い相関の存在があるだけ（一般にはp値0.05以下か否かで判定する。）では足りない。ばらつきも少なく、また、奏効率の増加に伴って大きな生存期間中央値の増加が見込めること（相関直線の傾きの大ききで判定）が必須となる。したがって、ここに示されるような相関に過ぎなければ、「妥当性」の前提となるべき相関の証明としては、極めて不十分なものとなる。

西乙H38の2＝東甲F11の2の記述を見ても、「RR値と生存期間中央値間に有意な相関が認められた」（p 4 左段、p 5 左段）としながらも、「RR値（奏効率）と生存期間中央値間に乖離が見られる。」として、相関について否定的な評価を前提とした論述を行っている。さらに、この論文の「考察」においても同様に、相関に関する否定的な評価を前提に、その原因について詳細に論述する（p 5 左段以下）。具体的には、各試験における測

定の不正確さ、選択バイアスの問題、上述した奏効率測定の限界、などといった点が指摘されているもので、これらの点からしても、この論文において確かめられたとされる相関を過大評価できないことは、同論文自体が自認しているものである。

以上のとおり、この論文においては奏効率と生存期間中央値との相関が確かめられたとされるものの、奏効率が延命効果に対する代替指標としての「妥当性」を有することなど何ら認められず、そればかりか、相関についてすら過大評価できないものである。これらのことは、この論文自体が認めるものである。

6 II 相承認の制度設計と、奏効率の位置づけ

抗がん剤開発、承認制度の枠組みについても、奏効率による延命予測に限界があることを当然の前提としており、安易に有効性を肯定することを未然に防ぐように考えられていたものである。仮に、西條、福岡両証人の述べるような「延命の蓋然性」が高い精度で認められるというのであれば、第II相承認制度を設計するにあたり、莫大な費用、時間、医療機関の負担を要する第III相試験を省略できる制度設計もあり得たはずである。しかし、実際にはそのように考えられたことは全くなかった。

(1) 「臨床試験計画（プロトコール）の作成と実施、並びに結果の統計解析とその評価について」（西甲D34＝東甲H22）の記載

旧ガイドラインが作成される2年前、厚生省がん研究助成金指定研究「固形癌の集学的治療との研究」班により、「臨床試験計画（プロトコール）の作成と実施、並びに結果の統計解析とその評価について」と題される論文（西甲D34＝東甲H22）がまとめられた。執筆者を見ると、西條証人や下山氏を初め、後に旧ガイドライン作成メンバーとなった者も数名含まれている。

この文書を見ると、との記載で、抗がん剤の臨床試験においても、第Ⅱ相試験が「探索」に位置づけられることを明確にしたうえで(p 2511右段)、

「小規模な第Ⅱ相試験的な臨床試験で、既存の治療法より**ずば抜けて良い治療効果を示すことが明らかになった場合は、中規模の第Ⅱ相試験は必ずしも必要ではない。むしろ、生存効果が得られるかどうかを目的にした第Ⅲ相試験的な無作為化比較試験に進む方がよい。**」

とする。この記載から、奏効率からすれば「ずば抜けて良い治療効果」で示した薬剤でさえも、なお、第Ⅲ相試験の省略はできないことがわかる。

このように、抗がん剤承認においては、いかに高い奏効率が見られたとしてもそれだけで市販承認の根拠としては足りず、単に第Ⅲ相試験に進めてよいというだけ、という位置づけとされてきたことがわかる。したがって、イレッサのIDEAL試験程度の奏効率の存在をもって、有効性が確実視できたなどという被告らの主張は失当である。

(2) 新医薬品課審査官（当時）による旧ガイドラインの解説

当時の厚生省薬務局新医薬品課審査官である川原章氏は、平成3年3月の「新医薬品研究開発フォーラム」という、旧ガイドラインの内容の解説（西甲D35=東甲H19）において、

「第Ⅲ相に関して特に申せば、これについては、**かならず延命効果を見る、ということである。腫瘍は小さくなったけれども、本当に命が延びたのかどうかという問題である。…外国の論文等の中には、腫瘍は縮小したけれども、延命については疑問な結果が出たといった報告も見られている（p 47）。**」

という腫瘍縮小効果による延命予測の限界を確認する。このため、旧ガイドラインによっても、原則的には

「第Ⅲ相は申請の時点までにできていれば一番よろしい。（p 48）」

と述べる。このように、旧ガイドラインにおいても、他の医薬品と同様、最終的には、第Ⅲ相試験において真のエンドポイントである延命効果で検証したうえで、承認が行われるべきという原則が維持されていたことは明らかである。

そのうえで、例外的に、Ⅱ相承認が要求される例として、

「腫瘍縮小効果が非常に高い薬、たとえば従来P R まで含めて20%程度までしかなかった癌腫に70%といったものが出てきたような場合に…そういった効果の高い薬もあるということで第Ⅱ相までで申請は認める。(p 48)」

と述べている。この記載から、「非常に」高い奏効率がなければⅡ相承認を与えない趣旨であることも十分に読み取れるものである。そうでなくとも、少なくとも既存の薬剤と同程度の奏効率しか示せていないような場面においてまで、あえてⅡ相承認を与えるようなことは許容しない趣旨であることが明らかである。

7 小括

以上、本項においては、高い奏効率が生存期間中央値の延長には繋がらない実例が、承認当時、既に数多く存在していたこと（前述「2」）、そもそも「奏効率」の判定方法は、第Ⅱ相段階におけるスクリーニング目的に沿うように定義されていて、延命予測の精度を最重視するような設定とはなっていないこと（前述「3」）、被告ら申請証人の論拠は、科学的な根拠を欠くものであること（前述「4」「5」）について論述を行った。これらの点からも明らかとなり、イレッサの評価にあたり、有効性の根拠として奏効率を用いるときに、これを過大評価してはならないものである。

第4 IDEAL 1, 2の奏効率から有効性を推測することの誤り

1 序論

イレッサ承認にあたっては、IDEAL-1, 2各試験で確認された奏効率を根拠に有効性が肯定された。しかし、一般に抗がん剤の有効性を推定するためには、最低でも、20%の奏効率の確認は必要と考えられているところ（後述「2」）、IDEAL-1, 2各試験を全体として見れば、このような最低限度の奏効率ですら確認されていなかったものである（後述「3」）。

被告らはドセタキセル試験と比較することにより、イレッサ奏効率を肯定的に評価すべきと主張するが、このような評価手法には歴史的対照による問題があり、実際にも単純比較を許さないような背景因子の違いが明らかとなっていたから（後述「4」）、イレッサ有効性を根拠づけることはできない。

また、他の既承認薬によっても、イレッサ承認当時、既に、IDEAL-1, 2各試験に見られたものを上回る奏効率が確認されていたものであるから（後述「5」）、被告らの前提は、イレッサの有効性を過大に評価するもので失当である。

2 抗がん剤の奏効率の確認に用いられる一般的基準

(1) 20%の奏効率を必要とする西條証言の内容

主尋問で、西條証人は、セカンドラインの抗がん剤の奏効率について、「20%以上であれば」ようやく「ある程度効く抗癌剤というところに分類」に達すると述べる（西乙E19＝西條証人主尋問調書p15～16、西乙E20＝西條証人主尋問調書p68）。さらに、平成14年当時に抗がん剤として市場が求めていた薬剤を問われ、

「先ほど、奏効率が30%と言いましたけれども、これは不十分ですから、これを50%くらいに持って行けるようなサイトトキシック

(細胞毒性)ドラッグです。(西乙E 1 9 =西條証人主尋問調書p 1 6)」

と述べている。ここから、まず、30%もの奏効率であっても不十分と断じており、さらにこの奏効率を20%も引き上げて50%に達することを目標としていることが読み取れる。したがって、西條証人も、20%という数値は効果の存在を予測しうる最低水準と認識していることが窺える。

この点に関しては、西條証人自身の論文で、統計解析におけるカットオフポイントを奏効率20%に設定した根拠の説明において、

「20%以上の奏効率を有する薬剤が非小細胞肺癌に対して活性あり、と考えられているからである。(西乙H 3 8 の2 =東乙F 1 1 の2 訳文p 3 左段)」

と述べている。ここで、「活性」とは「生物学的活性」とも呼ばれるもので、一般に、臨床的な真のベネフィットの確認の前段階として、客観的に抗腫瘍効果など、薬剤の作用が外形的に確認しうるかどうかを問題とする概念である。一般的には、活性すら見られない被験薬であれば、真のベネフィットは到底期待し得ないものと評価され、開発を中止すべきとされていた。

なお、西條証人は反対尋問になると、上記主尋問内容の確認の質問にも関わらず、この点を曖昧にした。その上で、IDEAL 1 結果である15.5%でも十分などとも反論もしている(西乙E 2 0 =西條証人反対尋問調書p 6 9)。しかし、これらの証言は、上記の証言や文献において自ら設定した前提を、都合良く場当たりの変遷させるもので、信用できない。

(2) 旧ガイドラインにおける期待有効率20%の記載

西條証人の上記評価は、旧ガイドライン(西乙D 7 =東乙H 7)における期待有効率20%の記載とも一貫する(西乙E 1 9 =東西條主尋問調書p 2 4でも確認)。

具体的には、旧ガイドラインにおいては、前期第Ⅱ相試験における期待有効率の設定に関して、一般的に20%以上を目標とすることとし、「この期待有効率以上の効果がなければ有用な抗悪性腫瘍薬としては認められないことになる。」との前提に立つ（西乙D7=東乙H7・7枚目）。

旧ガイドライン作成班の一員でもある下山正徳氏による旧ガイドライン解説においても、「一般的に有効率は20%以上を目標にする。どのくらい効く薬を開発するかが念頭になれば、新薬の開発はできないので、その目標を設定しているわけである。期待する有効率を期待有効率といい、固形癌では一般的に20%以上が目標となる…。この期待有効率とそのrejection error（すなわちβエラー）を設定することにより期待有効率を示さない薬、すなわち無効な薬を早期に判別する。」とされるべきことが明記されている。（西甲D15=東甲H10p126左段以下）

（3） 旧ガイドラインが、期待有効率20%を示した趣旨

もともと、旧ガイドラインが、一般的には、期待有効率を20%に設定すべきとしたのは、20%の奏効率が确实視される薬剤には当然にⅡ相承認を与えて良いとする趣旨などと解すべきではない。

なぜなら、上述のとおり旧ガイドラインは、原則形態としては第Ⅲ相試験まで終えて承認を与えるべきことを想定していて、20%の奏効率は、この原則形態において、第Ⅲ相に進むための最低基準を示したものに過ぎない（この点については、第Ⅲ相終了後の承認を前提とした通常形態の治験の進め方を論述した厚生省作業班による文献（西甲D34=東甲H22）においても、同じ期待有効率が設定されていることからわかる。）。そして、単に第Ⅲ相に進めても良いかという場面においては、未だ市販の前段階にとどまるため、有効性の判定をやや緩和することも許容されるのに対し、第Ⅱ相承認を与える場面においては、市販により多数の一般患者に流通する以上、より厳

格な基準を用いるべきだからである。

この点、第Ⅱ相のスクリーニング判定の際には、なるべく、有効なものを誤って振り落としたくないという要請（ β エラーの最小化をより重視する。）があり、これに沿うように期待有効率も想定される。20%という奏効率は、確実に効果が期待できるようなものではなく、西條証人も「まあ効くと判定するようなパーセンテージ」（西乙E 19＝西條証人主尋問調書25）といった程度の、あいまいで低水準なものではあるが、第Ⅲ相への移行の判定という目的の限りにおいては、これで十分と考えられたものである。

他方、Ⅱ相承認に求められる水準を考えてみると、上述の川原氏の論述（西甲D 35＝東甲H 19）を見る限り、かろうじて第Ⅲ相に進ませてやってもいいか、といった程度の被験薬全てにつき、医薬品承認を与える趣旨とは到底解されない。したがって、旧ガイドラインにおける期待有効率20%という設定は、単に開発中止とならないために最低限クリアすべき水準である、と理解すべきものである。

（４） セカンドライン以降の治療薬としての評価

この外、西條証人は、イレッサの低い奏効率でも許容すべき論拠として、非小細胞肺癌、特にセカンドライン以降の治療全般において、総じて、抗がん剤の効果が小さく奏効率も低かったことを挙げる（西乙E 19＝西條証人主尋問調書 p 25 参照）。

しかし、この立論は、まずもって前提に誤りがある。この当時、他剤による非小細胞肺癌のセカンドライン患者を対象とした試験における奏効率の結果を見ると、イレッサ承認当時であっても、併用で、33%（西甲H 48＝東甲F 82）、40%（西甲H 49＝東甲F 83）、単剤でも、19%（西甲H 50＝東甲84）、21%（西甲H 51＝東甲F 85）、5.2% CRを含む34%（甲H 52＝東甲F 86）などが報告されていた。

I D E A L 1, 2 の試験の計画にあたって、このような実態を反映してか、期待有効率を 15～20%, 閾値有効率 5% として被験者数が設定されている。I D E A L - 2 が期待有効率を若干低めの 15% に取っていることを除けば、旧ガイドラインに示された一般的な基準を正にそのまま採用しているものである。したがって、試験計画時においては、非小細胞肺癌のセカンドライン治療についても、通常の一般的な固形癌に求められる有効率の基準に沿うべきことが想定されていたことがわかる。

また、そもそも他剤の奏効率が軒並み低かったとしても、II 相承認のハードルを下げる論拠とはならない。なぜなら、奏効率は、真のベネフィットである延命効果を予測して初めて意味があるところ、他剤の奏効率が小さいからといって、論理的に、低い奏効率の被験薬によっても延命を期待できる確率が高まることにはならないからである。

(5) 小括

以上より、一般に、抗がん剤に求められる最低限の生物学的活性として、約 20% の奏効率が必要と考えられていたことがわかる。この点については、西條証人自身も認め、旧ガイドラインも同様の前提に立つものである。しかも、この水準は、第 III 相試験に進むために最低限必要とされるものに過ぎず、市販承認に要求されるような、より確実な延命予測の水準には達しないものとして理解すべきものとなる。

このような理解を前提として、以下、イレッサの I D E A L 試験において見られた奏効率について検討する。

3 I D E A L 各試験におけるイレッサ奏効率の評価

(1) I D E A L - 1, 2 において見られた奏効率の概観

これらの試験で見られた奏効率をまとめたものが、下の表である（西丙 C

1 = 東丙D1 p 470, 502)。IDEAL-1に関しては、申請資料概要に複数の表が掲載されているが、このうち、最も信頼性が高い数値は審査センター判定による表ト-73（西丙C1 = 東丙D1 p 470）であり、下の表や、後の論述においても、これらの数字を前提とすることとする。

試験	用量	人種	奏効率	信頼区間
IDEAL-1	250mg	日本人	25.5%	14.3~39.6%
		日本人以外	5.8%	1.2~15.9%
		合計	15.5%	
	500mg	日本人	25.5%	14.3~39.6%
		日本人以外	9.3%	3.1~20.3%
		合計	16.3%	
IDEAL-2	250mg	日本人以外	11.8%	6.2~19.7%
	500mg		8.8%	4.3~15.5%

(2) IDEAL-1の各群全体としての評価

上述のとおり、一般に、抗がん剤に最低限求められる奏効率は20%とされているところ、イレッサのIDEAL-1, 2のうち、平均的に高い数値を示したIDEAL-1の奏効率でさえ、承認用量の250ミリ群で全体として見れば15.5%にとどまっていたものである。しかも、この15.5%という数値は日本人群の奏効率によって大幅に引き上げられていたものであるが、西條証人も認めるとおり、日本人群の奏効率は、単に背景因子の偏りを調整すれば、外国人群の低い数値並みとなっていた可能性もあるという（西

乙E20＝西條証人反対尋問調書p71～72)。にもかかわらず、最低水準の20%もクリアできないというのであるから、この試験をもって、イレッサの奏効率を高いなどと評価できないことは明らかというべきである。

(3) IDEAL-2の各群全体としての評価

IDEAL-2についても、いずれの群においても奏効率は、わずか11.8% (250mg)、8.8% (500mg群)であり、一般に求められる20%奏効率の水準を大きく割り込む結果となっていた。したがって、この試験を前提としても、イレッサの奏効率を高いとは、到底評価しえないものであった。

(4) プロトコールに照らした評価

ア プロトコールの「解析方法」により有効性判定を行う必要性

以上、一般的な抗がん剤について設定される期待有効率20%を前提に、イレッサの奏効率の評価を行ってきた。もっとも、本来、イレッサの承認根拠となったIDEAL-1, 2の各試験の評価にあたり、まずもって、第一義的な評価基準として検討されるべきは、事前に、試験計画書(プロトコール)において設定されていた有効性の検定方法である。このように、プロトコールにおいて、事前に評価方法を明記しておくのは、治験の報告者が、後付けの、医学的にもっともらしい理由を付けて過大に有効性を報告することを防止するためである。つまり、ICH-E9にも示されたとおり、事前の予想に反して芳しい結果が得られなかった場合に、バイアスの影響や、不当な動機からごまかしてしまうことのないように、客観的に妥当な基準を事前に明示しておくものがプロトコールなのである(西甲P15＝東甲H3p23のV5.1, p28の5.5参照)。

そこで、原告代理人らにおいて、改めてイレッサのIDEAL1, 2の

試験結果を検討したところ、実際には、各試験のプロトコールに照らして見れば、複数の群において有効性の検定に失敗し、イレッサの奏効率が高いなどとは、到底評価しえないものであった。

イ プロトコールの「解析方法」における有効性検定の方法

I D E A L各試験のプロトコールにおいては、以下のような方法で、有効性検定を行うことが、事前に明記されていた。

I D E A L 1については、申請資料概要で、「奏効率の95%信頼区間の下限が5%を上回っていた場合、真の奏効率は5%以上であると結論づける。」（丙C1 p 462・表ト-66）と記載されている。

同様に、I D E A L 2についても、申請資料概要において、奏効率「5%は他に有効な治療がない場合の実薬の許容される最小率として選択される。」ことを前提として、試験結果から得られた奏効率の信頼区間下限がその5%を下回るようであれば、有効性はないものと評価すべきとされていた（西丙C1＝東丙D1 p 498「解析方法」）。

ウ 旧ガイドラインにおける閾値有効率との関係

このように、I D E A L各試験において、一般に求められる期待有効率20%を大幅に下回る、5%奏効率を基準としているのは、旧ガイドラインにも記載された一般的な閾値有効率の水準をもって、無効な薬剤の判定を行おうとしたためと考えられる。

閾値有効率とは、有効率の判定にあたって帰無仮説として設定される奏効率の値であるところ、下山正徳氏による旧ガイドライン解説によれば、

「近代的な方法では、期待有効率の他に、さらに無効な抗がん剤の判定に使う閾値有効率を適切に決める。この閾値有効率は5%以上とするのが通例である。」

とされる（西甲D15=東甲H10 p126左段）。

この点、当然のことながら、あまりに奏効率が低い抗がん剤については、到底、延命効果など期待できないもので、有効性は否定される。そこで、第Ⅱ相試験の症例数の算出にあたっては、一定の奏効率、例えば5%という目安を設定し、これ以下であれば開発中止としてよい、という前提で行われる。ここでの目安となる奏効率（上述の5%）が、「閾値有効率」と呼ばれるものなのである。

これを更に説明すれば、次のとおりとなる。例えば、ある被験者群で得られた抗がん剤の奏効率について、その信頼区間の下限が5%を割ってしまったとすれば、確率論として、真実の奏効率が5%以下である可能性、すなわち効果のない抗がん剤である可能性が一定程度以上に残ることとなる。統計学的に言えば、当該抗がん剤が無効な薬剤である可能性（帰無仮説）を排除（棄却）することをもって、当該抗がん剤の効果を証明しようとしたところ、その証明に失敗したということである。そのような証明ができない場合には、期待有効率以上の有効性を得る可能性はないとみなすことにする、という意味になる。この場合、市販など許されないのはもちろんのこと、その時点で開発を中止してよい、という結論になる。

なお、旧ガイドラインの閾値有効率の記述箇所において想定された場面は、より探索的な色合いの強い前期第Ⅱ相試験のものであり、その閾値は極めて緩やかな水準として設定されている。つまり、この水準は検定に失敗すれば開発中止と判定する、という意味での最低限度の基準なのであり、この検定に成功したとしても、延命効果の予測ができて市販承認が許容されるような帰結とはならないことについては十分に留意すべきである。

エ IDEAL各試験の結果

まず、IDEAL1を見ると、250mg海外群における奏効率の信頼

区間の下限（1.2%）においても、500mg海外群の同下限（3.1%）とともに、この5%を下回ってしまっている（申請資料概要にも同様の否定的な評価の記載がある（西丙C1＝東丙D1 p487～488）。）。

I D E A L 2 についても、500mg群の奏効率の信頼区間下限（4.3%）がこの5%を下回ってしまっている。250mg群の信頼区間下限は6.2%であるが、これは、プロトコールで最低ラインとされた11例をわずかに1例上回ってかろうじてクリアしたという結果であった。

このように、I D E A L 各試験においては、複数の群において、開発中止の結論を導くべきとされるような、低い水準における有効性の検定にすら失敗したという有様であった。後述のとおり、検定に成功した日本人群についても、背景因子の問題があった以上、全体としてイレッサの奏効率が高いなどと評価することは許されないものである。当然のことながら、延命効果が精度高く推認できるかのような結論を導くことはできない。

（5） 日本人群の結果について

この点、被告らは、I D E A L 1 の日本人群の奏効率によりイレッサの効果を強調するような主張をしている。

しかしながら、I D E A L 1 試験においては、日本人群と日本人以外群との背景因子に著しい違いがあった。特に、試験結果に大きく影響する患者の全身状態（P S）については、日本人群において、全身状態が不良の患者（P S 2）の割合が著しく少ないという偏りがあった。被告会社自身、I D E A L 1 での国内外の結果の差について最大の原因がこのP Sの偏りにあったことを説明していた。また、この患者群の背景因子の偏りを調整すれば、日本人群の結果も、外国人群の低い数値に近づいていた可能性もあることは、被告側申請にかかる西條証人ですら認めていたのである（西乙E20＝西條証人反対尋問調書p71～72）。

このようなことを考えれば、IDEAL 1 試験のうち、日本人群の結果のみを取り出し、患者背景の著しい偏りを無視して高い有効性が期待できるなどとするは全くの誤りである。

(6) まとめ

以上のとおり、IDEAL 1, 2 試験ともに、通常、抗がん剤に最低限要求される奏効率 20% の水準を達成することができず、また、プロトコールに予め明記された検定にすら失敗してしまった群も続出したのであった。ここで留意すべきは、プロトコールに規定された検定的水準は、これに失敗すれば開発中止としてもよいような、極めて緩やかな有効性の検定であったということである。イレッサに関しては複数の群で、この緩やかな検定にすら失敗したということであり、その有効性を過大評価することが許されないことは明らかである。

IDEAL 試験結果については、イレッサには臨床的に意味がありうる程度の腫瘍縮小効果は認められなかったもので、当然の帰結として、延命効果など期待すらできないものと判断すべきものであった。

4 ドセタキセル試験との比較が有する問題点 (背景因子の問題)

(1) 被告らの主張

申請資料概要 (西丙 C 1 = 東丙 D 1) において、被告会社は、約 10% の IDEAL-1 の海外奏効率も含め「いずれの民族群においても単独両方で臨床的に意味のある奏効率を得られ」とし (西丙 C 1 = 東丙 D 1 p 509), 審査報告書 (西乙 B 4 の 1 = 東乙 B 4 の 1) でも約 11.8% の IDEAL-2 の奏効率について同様の評価をしている。

このような評価の前提として、まず、セカンドライン非小細胞肺癌患者に対する延命効果を検証した Shepherd (シェファード) 氏によるドセタキセル

試験があり、I D E A Lの各群の奏効率がこの試験のドセタキセル奏効率7.1%（西丙H 2 2の2=東甲F 2 0）を上回ったことが主要な根拠とされている（西丙C 1 = 東丙D 1 p 5 0 9，西乙E 2 0 = 西條証人反対尋問調書 p 7 2）。

（2） プロトコールには反映されていない議論であること

しかし、そもそも、真にこのような評価を行うべきであったならば、上述のとおり、当初のプロトコール作成段階において目標症例数の算出にあたり、期待有効率、閾値有効率を低く設定すべきものであるし、これに応じた有効性検定の基準が事前に設定されるべきである（西甲P 1 5 = 東甲H 3 p 2 3 V 5. 1，p 2 8の5. 5。結果報告についてp 3 3参照。）。

にもかかわらず、I D E A L各試験の症例数設定の際には、何らこれに類する検討をしないまま、通常どおりの期待有効率（I D E A L 1で20%，2で15%）と閾値有効率（I D E A L 1，2ともに5%）が採用されていた。そして、有効性の検定においても、5%の水準が採用された点については、上述のとおりである。これらの試験計画を見ると、イレッサの奏効率が低いとの試験結果が出てから、慌ててドセタキセル比較を持ち出した議論がなされるようになったという経緯が読み取れる。

このため、ドセタキセルとの比較の議論は、その信用性について、極めて慎重に吟味すべきこととなる。

（3） 歴史的対照の問題性

このような後付けのドセタキセル比較を持ち出した経緯に目をつぶったとしても、ドセタキセル試験との比較においてイレッサの有効性を基礎づける論法は、歴史的対照を用いる問題性、つまり過去の試験結果との単純比較は誤った結論を導く可能性があるという問題をはらんでいる（西甲G 4 = 東甲L

49 p 211, 西乙D7=東乙H7・8枚目（西福岡証人反対尋問調書=東丙G58 p 37で確認））。

西條証人も執筆者として名を連ねる厚生省研究班による文献（西甲D34）においては、端的に、「RCTを敬遠して、歴史的対照などの非無作為化臨床試験だけに依存すれば、臨床医学の科学性が崩れ去るのは自明である。」とまで述べられている。この点に関連して、福岡証人も、再主尋問において「ヒストリカルな比較」（歴史対照による比較と同義）については問題があり、最終的には、第Ⅲ相試験による確認が必要と認めた（西福岡証人反対尋問調書=東丙G58 p 80-81）。

（４） PS-2患者の割合と奏効率への影響

この外、ドセタキセル試験との比較論は、以下のようなPS（パフォーマンス・ステータス）による患者背景の相違の問題点もある。

まず前提として確認しておくべき事項として、IDEAL-1の国内群と海外群の間で見られた約20%もの奏効率の違いについてさえも、統計解析（多変量解析）の結果、民族差などは考慮することなく、単に、各群の間のPS、腺がんなどの患者背景因子の違いのみで説明しうるものとされている（申請資料概要（西丙C1=東丙D1 p 493-494）、審査報告書（西乙B4の1=東乙B4の1 p 39））。このうち、全身状態が不良なPSが2の患者割合の違いを見ると、国内群は海外群の約2分の1（割合の差としては約8%）という違いがあった。

そこで、比較対象となったドセタキセル試験（西丙H22の2=東甲F20）の患者の背景因子を検討してみると、IDEAL-1の日本人250ミリ群では、状態の悪いPS-2の患者割合がシェファード試験の僅か4分の1程度しかなく、患者背景に大きな違いがあることが判明した（西丙H22の2=東甲F20 p 5の表1, 西丙C1=東丙D1 p 468の表ト-70）。西條証

人も認めるとおり，このような患者背景の違いを無視して，単純に結果を比較して論じることは妥当ではない（西乙E20＝西條証人反対尋問調書p72）。

そもそも，Shepherd試験については，原著論文中に，セカンドライン，すなわち

「プラチナ製剤で治療した非小細胞肺癌患者を対象にドセタキセル単剤療法を検討した7つの第II相試験が報告されている。これらの7試験では，…全体の奏効率は14～24%であった。（2頁）」

「我々が実施した試験の全体の奏効率7%は，どのドセタキセル第II相試験で報告された値よりも小さいが，…。このいくらか残念な結果は，より進行した状態の悪い癌患者を選択したことが原因かもしれない。（12頁）」

と記載があるとおおり，患者背景の偏りによって，通常期待しうるよりも低い奏効率となっていた可能性が明記されている。したがって，一般的な奏効率の比較対象として適切であったのか，疑問が残るものである。

さらに，IDEAL-1の海外群については，PS-2割合でいけば18.9%で，上記ドセタキセル試験のPS-2割合25.5%を下回り，これだけを見れば本来，より高い奏効率が期待できたはずであった。にもかかわらず，承認用量250mg群における奏効率は5.8%であって，ドセ試験の7.1%すらも下回る結果であった。（西丙C1＝東丙D1 p468「表ト-70」）。

(5) 西條証言も患者背景の問題を認めていること

患者背景に関連して，IDEAL-1の日本人群の奏効率が高いとの評価について再確認すると，IDEAL-1の日本人以外の250ミリ群でのPS2は18.9%（10例），日本人は5.9%（3例）と3倍以上の違いが

ある（西丙C 1=東丙D 1 p 4 6 8）。したがって、西條証人も認めるとおり、海外群並みの比率でP Sの悪い患者をグループに入れた場合には、日本人の群でも、I D E A L 1の海外群のような結果に近づく可能性がある（西乙E 2 0 =西條証人反対尋問調書 p 7 1 ~ 7 2）。I D E A L 1の日本人群の奏効率は、偏った状態の良い患者群での結果といえ、イレッサが日本人に対しては高い腫瘍縮小効果があるなどと即断することはできないものであった。

5 他の既承認薬のセカンドライン患者に対する効果

このように、Shepherdによるドセタキセル試験（西丙H 2 2の2=東甲F 2 0）との対比においても、イレッサの奏効率は必ずしも高い奏効率を示せていない。それでは、I D E A L - 1試験の全体としての平均奏効率15.5%は、他の既存の承認薬との対比において、ずば抜けた奏効率を示すものであったといえるのであろうか。

他剤による非小細胞肺癌のセカンドライン患者を対象とした試験における奏効率の結果を見ると、イレッサ承認当時であっても、併用で、33%（西甲H 4 8=東甲8 2）、40%（西甲H 4 9=東甲8 3）、単剤でも、19%（西甲H 5 0=東甲8 4）、21%（西甲H 5 1=東甲8 5）、5.2%CRを含む34%（甲H 5 2=東甲8 6）などが報告されていた。したがって、日本人群の背景因子の偏りで膨らんだ15.5%の数値を前提としてもなお、イレッサI D E A L - 1の奏効率は、既に患者の選択肢となっていた既承認の薬剤（もしくはその組み合わせ）以上の効果を示すものとは評価しえないものであるし、非常に効果が高いとか、ずば抜けているなどとは到底言い得ないような、お粗末なものであった。

以上より、イレッサについて、非小細胞肺癌のセカンドライン患者に限定したとしても、他の既承認薬を超えるような奏効率があったとは評価しえない。これは、イレッサを第II相で承認したところで、奏効率で見ても、患者に対し

て既承認薬を超える利益をもたらすものではなかった。したがって、延命効果の検証はもちろんのこと、副作用の確認も不十分なまま、あえて承認を急ぐべき必要性があったとは評価できない。

6 小括

以上より、IDEAL-1, 2各試験で確認された奏効率は、全体として見れば、一般に、抗がん剤の開発継続に必要な最低基準とされる20%の奏効率（前述「2」）ですら下回るものであり（前述「3」）、また、ドセタキセルのShepherd試験との比較も背景因子の偏りにより適切な分析とはいえないもので（前述「4」）、イレッサ有効性を根拠づけることはできない。さらに、他の既承認薬との比較においてもイレッサの奏効率が殊更、優れているともいえない（後述「5」）。

ゆえに、奏効率を基準としても、イレッサの有効性は高いとはいえない状況にあったものである。

第5 IDEAL各群の生存期間中央値による有効性の推測について

1 被告らの主張の概要

被告らは、以下のIDEAL-1, 2試験の生存期間中央値（生存期間中央値）が臨床的に有意義なものであると評価できるもので（西乙B4の3=東乙B4の3p54）、これがイレッサ承認の根拠の一つとなりうると主張する。

IDEAL-1 250mg群 7.6ヶ月

IDEAL-1 500mg群 7.9ヶ月

IDEAL-2 250mg群 6.5ヶ月

IDEAL-2 500mg群 5.9ヶ月

2 対照群のない試験における生存期間中央値の評価

しかし、このような対照群のない試験における生存期間中央値というものは、比較試験による延命効果の検証とは全く性質が違うものである点に注意すべきである。つまり、延命効果というものは、適切な患者割り付けによって患者背景因子を均等化したうえで、対照群との比較において、はじめて確認しうるものである。この対照群との比較、という性質は、「延命効果」というエンドポイントを論ずるにあたり、欠くことのできない要素なのである。

これに対し、適切な比較対照のない試験の被験者群における生存期間中央値というものは、その群の患者背景因子によって大きく変動しうるものであり、仮に、一見良好な数値が得られたとしても、これが、被験薬の効果によって得られたものか、単に患者群の背景因子の偏りによってもたらされたものなのかを見分けることは不可能である。したがって、ある試験における生存期間中央値を、別個の試験や、一般的な臨床成績などと対比したところで、有効性の根拠としては全く信頼性がないと言っても過言ではない。

よって、上記の I D E A L 各試験の生存期間中央値をもって、イレッサの有効性を根拠づけることはできない。

3 生存期間中央値は副次的評価項目で過大評価してはならないこと

申請資料概要の試験計画から、I D E A L 各試験において、主要評価項目はあくまで奏効率であり、生存期間は、他の複数の副次的評価項目のうちの一つでしかないことが明らかである（西丙C 1=東丙D 1 p 4 6 2, p 4 9 8）。

したがって、平山証人も認めるとおり、承認当時、イレッサの有効性の主たる評価対象となるものはあくまで奏効率であり、生存期間中央値の数値は副次的なものとして捉えるべきものである。これは、I D E A L 各試験プロトコールにおいて、生存期間中央値は主要評価項目とされておらず、副次的な位置づけしか与えられていないことの当然の帰結である。

このように、複数の副次的評価項目がある場合に、そのうちの一つの指標で良好な結果を示したとしても、それは、偶然の要素が入り込みやすい外、単に「いいとこ取り」をしているだけの可能性を否定することができず、信頼できる有効性の根拠を示すものではない（西甲D 3 4=東甲H 2 2 p 2 5 1 1 右段「エンドポイントが多くなれば、本来差がないものでも、…もっとも大きな治療差が認められたエンドポイントを事後的に選択することで、判断を誤ることになる。」との記載参照）。

I D E A L 各試験の主要評価項目（プライマリーエンドポイント）は、あくまで奏効率なのであるから、その結果を最重視すべきは当然である。生存期間中央値を含め、他の指標の肯定的結果を過大評価してしまうことで、誤った結論に至る可能性がある点には注意が必要である。

4 既存薬における生存期間中央値の報告の概要

西甲H 4 7～5 2=東甲F 8 2～8 7は、イレッサのI D E A L 試験と同種の、非小細胞肺癌のセカンドライン第II相試験である。これらの試験における各群の生存期間中央値を見ていくと、それぞれ9ヶ月（西甲H 4 7=東甲8 7のArm A）、8ヶ月（西甲H 4 7=東甲8 2のArm B）、8.5ヶ月（甲H 4 8=東甲8 2）、42週間（甲H 5 1=東甲8 5）、40週間（甲H 5 2=東甲8 6）と、いずれもI D E A L 各試験の結果を上回る数値を示している（甲H 4 9, 5 0=東甲8 3, 8 4は、報告なし）。

5 審査報告書の生存期間中央値分析における比較対象

(1) I D E A L - 1 の評価について

なお、上述の「審査報告書（3）」（西乙B 4 の3=東乙B 4 の3 p 5 4）のI D E A L - 1 の生存期間中央値の評価にあたり、比較対象として用いられている「J Clin Oncol 18:2095-2103, 2000」という研究は、実は、奏効

率の比較対照ともなった上述のShepherdのドセタキセル試験である（西丙H 2 2の2=東甲F 2 0）。上の奏効率の確認で見たとおり、歴史対照の問題、患者の背景因子がIDEAL-1のものはPSが良好で単純な比較が許されない点については、奏効率の評価と同様である。特に、生存期間中央値については、患者背景因子によって、極めて大きな変動を見せるもので、異なる試験の間での比較はほとんど意味を有しない。ましてや、Shepherd試験とIDEAL各試験の比較においては、予後に大きく関わるPSや腺がん割合といった患者背景因子に大きな違いがあることは、一見して明らかである。にも関わらず、これを無視して単純比較しようとする被告らの態度は、科学性の放棄とさえも評価しうるものである。

特に、生存期間中央値の評価との関連で着目すべきは、腺がん患者が、一般に長期生存期間中央値を示すことである。例えば、あるセカンドライン非小細胞肺癌試験（西甲H 5 1=東甲F 8 5）においては、腺がん、非腺がんの生存期間中央値はそれぞれ5 1週、2 2週と、倍以上の差を示している。そこで、Shepherd試験の腺がん患者の割合を検討すべきこととなるが、この重要な患者背景について、論文中には明記されておらず（西丙H 2 2の2=東甲F 2 0）、生存期間中央値比較をするための前提情報を欠いているというべきである。

以上より、IDEAL-1の結果から、安易にイレッサの延命効果の存在について推認することは慎むべきである。

（２） IDEAL-2の評価について

IDEAL-2の生存期間中央値評価にあたっては、「プラチナ系抗癌剤及びタキサン系抗がん剤治療後」すなわちサードライン患者群において、「平均的な予後は4ヶ月程度と推測される」という前提事実に依拠しているのであるが（西乙B 4の3=東乙B 4の3 p 5 4）、その根拠はここに明示されて

おらず、全く不明確であり、そもそも歴史対照の評価にあたって不可欠となる背景因子や治療水準の推移等の影響について検討すらできない。したがって、IDEAL-2の生存期間中央値評価が「臨床的に有意義である」と断じるには、全くもって科学的根拠を有さないものである。

6 まとめ

以上より、IDEAL-1, 2の生存期間中央値をもって、イレッサの有効性の根拠とするのは、完全な誤りである。

第6 承認時点のイレッサの有効性評価についてのまとめ

これまで述べてきたように、そもそも、抗がん剤の第II相試験での奏効率をもって延命効果を推測すること自体に大きな問題性があることは、イレッサ承認時点では明らかとなっていた。

そして、イレッサについてIDEAL試験結果を具体的に見ても、その奏効率を積極評価することには様々な点において問題があった。

更に、対照群を置かないIDEAL試験での生存期間中央値などをもって、イレッサの効果を評価する根拠とすることは全く認められないことであった。

このような点を総合して考えれば、承認時点において、イレッサについて相当の腫瘍縮小効果があると評価すること自体に大きな問題があった。むしろ、IDEAL 1試験の日本人群には、外国人群や別のドセタキセル試験の患者群と比較しても、際だって全身状態の良好な患者に偏っていたという患者背景の偏りなどもふまえて試験結果を考えれば、イレッサが日本人の非小細胞肺癌患者の治療において有効性、すなわち延命効果を有しない薬剤である可能性を念頭に置くべきような状況にあった。