

## 癌の臨床試験における Survival Benefit の評価 —全治か延命か？—

中島聰總

癌研有明病院顧問

従来、癌の臨床試験における Survival benefit の評価には Generalized Wilcoxon,あるいは Logrank test など, non-parametric test がもっぱら使用されてきた。今回、本誌に掲載された医学統計解析グループ(代表:前谷俊三)は2つの胃癌補助化学療法の比較対照試験を引用しつつ、彼らの新たな指標による Survival benefit の評価を試みた。この度、私はあらかじめ本論文を review する機会を得たので、若干のコメントを記すことにした。

### 1. 引用された2論文に対する評価

まず著者らはオランダ胃癌グループの D1 vs D2 比較対照試験(RCT)のデータと解析方法について詳細に検討した。本試験の結果を受けて西欧諸国では胃癌の標準的治療として D1 surgery が定着した感がある。他方この試験では術後合併症、死亡率がわが国では信じられないほど高率で、日本人指導者の関与があったとはいえ、Technical compliance からみて D1 vs D2 の比較を可能にする土壌が未だ形成されていないとの見方が大勢を占めている。著者らはオランダグループの1999年と2004年の論文

の再発曲線を詳細に比較して、両論文には大きな矛盾があることを指摘した。すなわちオランダグループの複数の論文において D2 の Survival benefit についてはかなりニュアンスの異なる結論が述べられており、D1 vs D2 debate において結論が出たとは言い切れない。この点について著者らはこの論文に対する Debate を継続させるべきで、N Engl J Med または JCO に反論を投稿すべきであろう。

第2の引用論文はわが国で行われた補助化学療法の薬剤強度(Low vs high dose)を比較した臨床試験であり、わが国の臨床試験としては珍しく1410例の症例が集積された。全生存率において両群の間に差を認めなかったが、n(+)/ps(-)のsubsetにおいて再発を event とした生存曲線において有意差を認めた。これはJCOG8801 studyにおける結論と完全に照合するものであった。しかしこの試験が行われた当時(1987-1990)は胃癌治癒切除後における標準治療は「化学療法なし」であり、surgery alone を対照としない LD vs HD RCT の negative result は多くの関心と呼ばなかった。

## 2. 引用2論文に対する新しい評価

従来、臨床試験の Survival benefit は冒頭にも述べたとおり、Non-parametric test が行われてきた。著者らによればこれらの test は試験、対照両群の生存期間の順位検定であり、治療により全治した平均余命の差を検定しているものではないと批判している。著者らは新しい指標として平均余命を挙げ、Parametric model (Boag) により平均余命を比較することを推奨している。著者らによれば平均余命は生存曲線の下面積 (AUC) と定義した。生存曲線が0にいたるまで観察することはきわめて長期間を要するが、著者らは期間限定モデルにより長期間の観察を要しないで、平均余命の計算することを可能にした。Parametric model として Boag model が最適であるか否か、さらに期間限定モデルの妥当性については統計の専門家の検証が必要であろうと思われるが、臨床家にとって Survival benefit を評価する新しい尺度を得たことは極めて大きな意義があると思われる。

著者らはこの新しい尺度を用いて、D1 vs D2, Ld vs HD の両試験において、D2 および HD の有用性を結論付けた。

## 3. 新しい評価法の今後の問題

臨床試験の Survival benefit を評価する際、従来の NPA 法と新しい PA 法のいずれが臨床家にとって望ましいであろうか？従来の NPA 法は治療による一時的な延命も全治も区別されず、さらに観察期間中の早期における event を過大評価する欠点がある。他方、前述の欠点はあるものの、従来の NPA では early stopping rule を設けて「治療群、対照群のいずれを選択するか」を早期に決

定することが可能である。これに対して PA では信頼の置ける model の parameter を推定するためには疾患特異的生存曲線が水平線に近づくまで観察する必要があると思われるので、中間解析における stopping rule には採用しにくいのではないだろうか？感覚的には治療により全治した患者を対象とした評価、すなわち平均余命を指標とするほうがより適切に癌に対する治療の治癒力を評価できるのではないかと思われるが、NPA と PA とが相互に他を排除しあう関係にあるとも思われぬ。いずれの方法で検定して positive result である場合はまったく問題ないが、一方の方法では positive、他の方法では negative という marginal cases の場合はどうすべきであろうか？どちらか都合の良い結果を採用するというのでは無節操のそしりをまぬかれぬ。こうした非難を避けるためには臨床試験のプロトコールに最初からどの検定法を用いるか prospective に記述しておくことが必要であろう。

著者らは引用した2論文につき、PA法を用いて再検討し、D2とHDの有用性を導いた。著者らは引用論文の原著者とは独立した関係にあり、PAを用いた結論も原著者の結論とは独立した意味合いを持つであろう。しかし著者らの結論が原著者らの結論を否定することもできないと思われる。D1 vs D2 について「debate is over」という意見に反証することが著者らの結論の正当な位置づけではないだろうか？

平均余命の算定には疾患特異的生存率の算出が基礎にあり、follow upの際の死因の調査に正確性が求められることは論を待たない。

## 結論

臨床試験の Survival benefit を評価する尺度として平均余命が提唱されたことは今後の臨床試験の企画, 評価の段階に極めて画期的な影響を持つものと思われる。臨床家

の評価もそれだけ多面的となり, 評価の妥当性が増すものと思われる。今後, 実地の臨床試験において prospective な評価を重ねる必要がある。