

特別講演

主催 埼玉医科大学 国際医療センター トランスレーショナルリサーチセンター,

後援 埼玉医科大学 医学教育センター 卒後教育委員会

平成22年3月29日 於 埼玉医科大学 国際医療センター 2階会議室,

本部棟 1F 大学院講堂, 総合医療センター 第4会議室

がん医療開発における臨床統計学入門

竹内 正弘

[北里大学薬学部臨床医学 (臨床統計学・医薬開発学)]

日本国内で、優れた新薬を迅速に患者さんの元が届けるために、1998年8月に厚生省(現厚生労働省)より「外国臨床試験データを受け入れる際に考慮すべき民族的要因について」のガイドライン(E5ガイドライン)が通知された。このガイドラインを受けて、小規模の臨床試験、所謂ブリッジング試験を実施し、すでに海外で承認されている新薬を、既存の海外臨床試験データを日本に外挿することにより、新薬の承認申請が可能になった。この動向に合わせ、Drug Lag問題を解消するために、近年では国際共同臨床試験に積極的に参加し、有用な新薬の承認申請を各国と同時期に実施するという試みが始まってきている。このような環境において、実際に臨床試験に携わる医療従事者が正しく深く臨床試験を理解することは極めて重要である。そこで米国食品医薬品局(FDA)薬品評価研究センター(CDER)生物統計I部で長年勤務され、また厚生労働省薬事・食品衛生審議会専門委員を務められるなど、医療開発における臨床統計学の第一人者である竹内先生に臨床試験実施の目的、意義などを臨床統計の観点から解説していただいた。以下、そのご講演の概要である。

臨床試験では、まずターゲットとなる患者層があり、それに対して薬剤等を用いた治療を行い、その効果を科学的に評価するものです。例えばターゲット患者層を仮に非小細胞肺癌とし、初回治療として薬剤Xによる治療抵抗性を示した症例とします。この患者層から400例のサンプル抽出を行い、ある薬剤Aと薬剤B(標準薬)にそれぞれ200例ずつ割り付けるものとします。このA群とB群を公平に比較する妥当性を与えるものとして、A群・B群に影響を与える既知、あるいは未知の因子を両群に均一に割り振る無作為化が

重要です。特に臨床研究では評価項目に影響を与える因子、すなわちPS・性別・組織型・施設といった因子で作られた層別に割り付けを行う層別無作為化が良く行われております。

評価項目ですが、主要評価項目は臨床試験において検討したい薬効・有効性を示す項目で、これは客観的に評価できる測定指標でなければいけません。例えば延命効果は数値化が可能な客観的指標ですが、全般改善度は評価者間でのばらつきが考えられる主観的指標で、主要評価項目としては不適です。副次評価項目は、対象としている薬剤の特徴を表す指標で、例えば抗がん剤を使用した場合には、QOLや腫瘍縮小効果などが挙げられます。これらの項目について、有効性が数値化され、統計的に解析されます。

推定は抽出したサンプルよりターゲット患者層の有効性を推測することで、平均値や中央値といった点推定と、95%信頼区間で表します。これは母集団の平均値・中央値が95%の確立で入っている区間であり、臨床試験を100回行った場合、5回は真の値を外すことですので、通常我々は1つの臨床試験からある1つの信頼区間を偶然的に得ているに過ぎないということに注意を払う必要があります。

次に仮説検定ですが、これは得られた推定値の相違が果たして偶然的なものなのかそれとも本質的なものなのかを統計学的に考察する方法です。臨床試験での仮説としては、薬剤Aは薬剤B(標準薬)と比較して有効性がある、例えば12ヶ月に対して10ヶ月のような延命効果があるかといった優越性と、薬剤Aは薬剤B(標準薬)と比較して劣ってはいない(12ヶ月 VS 13ヶ月)という非劣勢の2つの仮説があります。優越性を例にとりますと、薬剤Aと薬剤Bの有効性は同じである

という帰無仮説と、それに対して薬剤Aと薬剤Bの有効性は相違している(両側検定)あるいは薬剤Aは薬剤Bと比較して有効である(片側検定)という対立仮説があります。検定の方法は、この薬剤Aと薬剤Bの有効性が同じであるという帰無仮説を正しいとして、臨床試験より推定された値(有効性の差・比)が起こりうる確立を計算し、その指標である検定統計量p値を計算します。このp値には第1種の過誤： α エラーと第2種の過誤： β エラーがあります。 α エラーは検定統計量より導かれ、帰無仮説が正しいと仮定したときに、A群とB群の有効性が相違している確率です。すなわち、 $p=0.05$ の場合は、同じような臨床試験を何回実施しても、20回に1回しか同様の薬効差が観測されないということです。すなわち、帰無仮説の下では、現実の薬効差が起こりにくいと結論でき、よって、帰無仮説を正しいと仮定したことが間違っており、この帰無仮説を否定する基準となります。次に β エラーですが、これは薬剤Aと薬剤Bの有効性が相違している(対立仮説が正しい)時に、薬効性が同じである(帰無仮説が正しい)と結論づけてしまう確立のことです。検出力は $1-\beta$ エラーで得られるもので、これは対立仮説が正しいときに、対立仮説を正しいと結論づける確立で、A群とB群に対して有効性がある時に、有効性があると結論づける確立です。

臨床研究のプロトコール作製時には、必ず症例数を

どのように設定するかと問われます。症例数を計算するためには、1) 予想される薬剤Aの薬剤Bに対する有効性・薬効、そして2) α エラーと β エラーの値、という2つの情報が必要です。例えば延命効果が12ヶ月(薬剤A)と8ヶ月(薬剤B)の相違とすると、4ヶ月の延命効果が得られるわけで、この情報が非常に重要となります。逆に、延命効果がわずか1週間でも、症例数を増やすことによってp値を小さくし帰無仮説を棄却できるかもしれませんが、果たして臨床的に意味があるか否かは別問題です。症例数設計の表を作成し、これを基にして、臨床研究者と統計学者の両者で十分な協議・検討を行い、最終的な目標症例数の設定を行います。臨床試験を計画する上で、十分な検出力を保證することが成功するための秘訣です。

以上をまとめますと、臨床研究・研究データ無しでは、より良い治療・薬剤は臨床現場に提供できません。統計学者は試験・研究の質の向上・管理のためのお手伝いを行います。最後に、チームでの共同作業・コミュニケーションが大切であり、現場の声を反映させることが、試験・研究デザインの品質管理に重要で、あらゆる角度からの検討が必要です。

ご講演後も、臨床研究について現在抱えられている問題点について聴衆より数多くの質疑がなされ、極めて有意義なご講演であった。

(文責 江口英孝, 西山正彦)