

第4回プラセボ対照試験に関する専門部会

日時 平成27年5月8日(金)

15:00~

場所 PMDA会議室21~24(14階)

< 開会 >

山本部会長 定刻を過ぎましたので、第4回プラセボ対照試験に関する専門部会を開催させていただきます。本年度になってから初めての部会ということになります。本日はお忙しい中、ご出席いただきましてありがとうございます。

まずは、事務局のほうから委員の出席状況の報告と資料の確認をお願いしたいのですが、事務局長が新しく代わりましたので、自己紹介を含めてお願いします。

< 出席状況確認及び配付資料確認 >

蛭田事務局長 ただいまご紹介いただきました蛭田と申します。4月1日付けで科学部委員会事務本部事務局長を拝命いたしました。今後ともどうぞよろしくお願いいいたします。

それでは、資料の説明等をさせていただきます。まず、委員の出席状況です。新委員会からご参加の委員も含めまして、16名の委員で構成されております。本日、12名のご出席予定でございます。現在、10名の委員の先生方に出席いただいておりますけれども、赤池委員と加藤委員が遅れて参加とご連絡いただいております。また、本日、部会長と事前にご相談させていただきまして、本日の議論と関係の深い、外部有識者といたしまして、日本製薬工業協会の稲垣治先生にご出席いただいております。後ほど稲垣先生にプレゼンをお願いし、その後の議論にも参加いただきたいと思います。

次に、配布資料の確認をいたします。まず、座席表、資料目録、議事次第、資料1、資料2と配布させていただいております。本日は全ての資料をお持ち帰りいただけるものです。事務局からは以上です。

山本部会長 ありがとうございます。委員の方は、名簿が最後に入っていますか。荒川先生が代わられたので、そのほかにも所属が代わられた方はまた事務局の方にお知らせいただけたらと思います。よろしくお願いたします。

それでは、議事に入ります。最初の議題は、前回に引き続いて、治験におけるプラセボ対照試験の現状についてです。前回は、プラセボ対照試験の現状、課題について、荒川先生にご紹介いただきましたが、今回は、製薬企業からのプレゼン、プラセボ対照試験の実施の現状と課題について、稲垣先生からご紹介をいただきたいと思います。どうぞよろしくお願いたします。

< 議題 1 : 治験におけるプラセボ対照試験の現状について >

稲垣氏

日本製薬工業協会医薬品評価委員会の稲垣です。本日はこのような機会を頂戴し、ありがとうございます。

本日は、企業におけるプラセボ対照試験の実施状況ということで、製薬協の中で調査を行いましたので、その結果についてご紹介をさせていただきます。最初に改めてご紹介させていただきますと、日本製薬工業協会とは、研究開発指向型の製薬企業 72 社が加盟している任意団体です。ただ、この 72 社ですが、自社の医薬品の開発の状況等については各種様々でして、毎年いくつもいくつもテーマを走らせているところから、数年に 1 回申請するというところまで各種ございます。モットーとしましては、患者参加型医療の実現というところで、製薬産業に共通する諸問題の解決等についての調査を行い、提言等を行うという活動を行っています。

その中で、私が所属しております、医薬品評価委員会と申しますのは、前臨床から臨床試験、そして製販後の調査というあたりの、いわゆる実際の患者さんのデータ等を集めて、その段階での問題等について検討する委員会です。今回、プラセボ対照試験ということはどういうものかということがありましたので、いろいろと話がありますが、最近の状況はどうなっているのかということで、改めて製薬協の企業の中で調査を行わせていただきました。お話いただいてから時間があまりなかったので、調査の対象としては、医薬品評価委員会運営委員会全部の 26 社なのですが、この委員会及び医薬品評価委員会の中にあります、臨床評価を担当している部会のメンバー等で、全部で 27 社の関係者にアンケート用紙を送り回答を依頼いたしました。試験といたしましては、最近の状況が知りたいなということで、2010 年 1 月以降に治験届を出した時点で、プラセボを対照とした試験についてどんな状況でしたかということで、回答をお願いいたしました。ただ、先ほどの会社によって治験の実施状況にかなり差がありますので、特定の会社の意見で動いてしまうというのもどうかと思いましたので、一応 1 社 5 試験ぐらいということでアンケートをお願いしております。その結果、27 社に送りまして、22 社より 65 試験の回答を受領いたしました。各社平均すると 3 試験ということになりますけれども、この中には 1 試験というところから 5 件というところまで、バラエティに富んでいます。内容は後ほどご紹介いたしますが、P1 試験が 2、P2 が 21、P2/3 として 7 つ、P3 が 35 というような形で計 65 になっております。試験の実施地域ですが、国際共同試験において日本でやったものというのが 24、国内での試験は、これが 40 で、回答として記

載なしというのが1でした。

プラセボ対照試験を実施した領域はというところなのですが、ここに
ありますように、やはり一番多かったのは、精神疾患領域とがんでして、
それぞれ25%、16%というところでした。ただ、精神領域、がん領域と
いうところで、これらの領域ではもっと多いのかなと予想していたので
すが、そう極端に差が出ているわけでもなくて、代謝・内分泌系領域で
も12%ぐらいありまして、予想していたところが多いものの、ずば抜けて
多いわけではない、ほかの領域でもプラセボ対照試験を一度はやって
いるのかなというのが今回の感想です。添付文書を出した趣旨としては、
ここにありますように、2010年以降に調査を行っておりますので、12年、
13年、14年、15年になってからさすがに少なかったのですけれども、大
体同じぐらいの感じで毎年プラセボ対照試験を行っております。したが
ってある程度正常状態にあるところでの調査内容という形になるかなと
思っております。

その下のところの、国際共同治験であるか否かというところで先ほど
のような話がありまして、国際共同が24試験37%、国内治験が40件、
63%ということです。対象として年齢はというところは、大部分が成人
かなと思っていたのですが、やはり12歳から18歳を対象にした試験も
11%、児童を対象にしても4%で、子どもさんを対象にしてのプラセボ試
験というのを行っておりました。ただ、後ほどご紹介しますが、苦労し
た治験だというような話も伺っております。投与期間ですが、12か月未
満とか、12月よりも長期というようなプラセボを投与する期間、領域に
もよるのでしょうかけれども、このような形のものもございました。では
実際どのような形で試験を行ったかということなのですが、やはり大き
くは2つで、まず、純粹に実薬とプラセボの対照試験、プラセボに対す
るヘッドツートヘッドを行ったというのが、グラフの黄緑のところでは
58%過半数がプラセボとの実薬の対照試験というところでもございました
が、基礎情報がありまして、その上にプラセボ、あるいは実薬を上乗せ
する上乗せ試験というものも37%。おそらく癌の領域ですと、この形が
多いと思うのですが、こういったものがありました。

それから、置換え試験ということで、治療を減らしてプラセボに置き
換えるという形の試験というのもありました。

意外だったのが、プラセボを入れて対照プラセボと、標準薬、比較対
照薬を入れた3群の試験というのは意外と少なく、3%というような結
果でした。

そして、下のグラフになりますが、プラセボ対照試験を実施した会社

はどういう工夫を行ったかというところなのですが、様々な工夫が行われておりますが、ランダム化、それから、もう 1 つは多かったのが、やはり何かあったときにレスキューしますよというところで、レスキュー薬の提供、あるいは何かあったときの対応を明確にすることによって、プラセボ対照試験の被検者、患者さんの組入れを促進するというような形があったかと思えます。

では、実際プラセボ対照試験を行ってどうだったかということなのですが、大部分の行った側は、価値はあったと見ておりまして、不明と 25%あるのですが、これは実は大部分が試験を継続中でして、結論が出ていないので、今の段階でよかったかどうかは判断できませんというような回答でした。少なくともプラセボ対照でやって後悔しているという回答はどこからもなかったというのが我々の調査の結果でした。ただ、今回の調査が、プラセボ対照試験を実施したところから回答を頂くという形で調査をしております、プラセボ対照試験を計画している、あるいはやりたかったがやらなかったというところの声は挙がってきておりません。その部分については、割り引いて考える必要はあるのかなとは思っておりますが、やったところはそれなりにやってよかったと思っているという印象を持っています。

実際やってみてどのような感想を持ったか、どのような苦勞をしたかということフリー形式で記載いただきました。まず、プラセボ対照試験ということ最初計画した段階でどういうことを心配したか、あるいはどういう問題があるというように考えたかというところなのですが、プラセボ効果をいかに少なくできるかというようなところで、CNS 領域という形での回答があったのですけれども、プラセボ効果がある程度あるような領域ですと、ここがあまり大きくなると判断しづらくなりますので、クリアな形になるようにするか、あるいは、実際の効果のある対照薬でも、プラセボに対する優越性を検出しにくいというような話がある中で、どのように検討して、クリアカットな効果を出す工夫というのが 1 つの声であったかと思えます。

同意がなかなか取りづらくなるのではないかと心配した。そのためにレスキュー薬を設定して、どういうときになったら、治験中止にするかという形、危ないときにどうしたら患者に納得してもらえるか、治験に参加していただけるかという手法をどう組むか。そのところが 1 つの苦勞であったかと思っております。ただ、ここに関しては、レスキュー薬を提供して、でも入っていただけないのではないかと、途中で辞めてしまう人が増えるのではないかと心配していたけれども、実際

ほとんど中止例はなかったというような、うまくコントロールして治験を実施できたところもありまして、このようなところはいろいろ工夫があるのかなと思っています。

もう一つ、既に適応外で、その当該薬剤が使用されているために、プラセボ対照試験という形で治験しようとしたとき、もう既にその薬は使っていますよということで、症例登録が難しかったというような話もありました。あとは、施設によっては、治験施設での責任医師の候補者の方に話をしたところ、実施可能性が低い、あるいはこれはどうかというようなことで断られたというケースもありました。

小児試験のほうで、患者さんのお子さん及び親は早急に症状が改善することを望んでいるため、数週間にわたるプラセボ投与でも、やはり難しいと言われたことがあります。あるいは、最初は問題なかったのですが、途中で治療薬が出てきて、途中からプラセボ対照試験が成立しなくなるかもしれないなど、はらはらしたという話もありまして、やはりまだプラセボ対照試験では、患者さんの同意取得、あるいは施設側からの抵抗が1つの懸念として残っています。

ということで、実際やってきたときに、そういうところで問題があったという話もありました。プラセボの場合、盲検化維持のための検査値のマスク期の設定とか、テクニカルな話もありますが、大部分は病気間によって患者に拒絶されたり、あるいは職業離脱、あるいは同意取得が取りづらかったりというような話での問題点でした。ただ、そうは言いましても、科学的な評価の点ではやるべきであったという判断をしています。では、やった施設に対して、どのような視点について、プラセボ対照で実施すべきかと考えていますかという質問に対しては、科学性の担保が必要な場合は、やはりプラセボ対照は必要でしょう。特に客観的な数字で示されない自覚症状がメインのエンドポイントになるような場合、あるいはプラセボは効果が高いと想定されるような治験の場合はいいのではないかとということです。

また、POC試験で正確に効果を判定したい試験という場合も、やはりプラセボ対照で行うべきでないか。さらには、可能な限り、原則としてプラセボ対照試験を実施すべきであるというような会社がありました。

以上、今回プラセボ対照試験を実施した会社からの意見をまとめた結論ですが、話を聞いた限り、試験を実施した会社では、科学的評価が必要、あるいは、プラセボ対照試験が必要と判断したならば、プラセボ対照試験を実施している。必要と思えばやっているというところだったかと思います。対照領域としては CNS 領域とがんが一番多かったというところ

ころですが、デザインとしては直接上乗せ試験で 95%、3 群が少なかった。そしてプラセボ対照試験ということで、実施に関して障害はありましたが、企業としてやれるのではないかと判断したものについては、何とかできる、実施可能なレベルであったというところで、必要な試験についてはやれるというようなことが、企業側で考えているというのが、今回の調査の結論でした。ということで、医薬品評価委員会としまして、やはりプラセボ対照試験というのは、必要な場合は実施したい。ただ、全部が全部プラセボ対照試験でと言われるのも厳しい。と言いますのは、同意取得等の問題がある。ただ、必要だと思ったときには我々としてはやりますというのが会社としての見解という形でまとめられるかなと考えています。以上です。

山本部長 貴重なアンケートの結果をありがとうございました。一つ一つについて議論したいと思いますので、稲垣先生のご発表に何かご質問、ご討議はありますか。全体でプラセボ対照試験が行われた、どこを全体とするかは難しいですが、それとできなかつたものは、どのぐらいの割合であるのでしょうか。

稲垣氏 実際、今回アンケートを最初 26 社に出して、22 社から回答が来ているというところで、残り 5 社が実施しなかつたというところですが、その中で実施しなかつた試験がどれぐらいあつたかという話かと思うのですが、それはデータが分かりません。ただ、20 何社中 22 社はこうやっていきますというところですので、プラセボ対照試験に対する抵抗感を持っていないというところで、意外とやる必要があると判断したら、やっているという状況かと推測しております。

山本部長 1 つの会社でもやったものだけは出してきたけれども、やっていないものがたくさんあるのか、それはどのぐらいですか。

稲垣氏 それぞれの会社の中で経験的にやれるもの、やれないものは中でノウハウを持っていて、やれないものは、特に患者が入っていただけない試験等はあるのだろうと。でも、そこはノウハウであり、ただ必要だと考えた試験については、今回やっていますという話で回答が来ていますので、必要だったらやっているのかと知っているところです。

佐藤委員 試験全体の中でプラセボ対照試験は、何割ぐらいやられているのですか。

稲垣氏 今回これで全体の数について調査しておりませんので、パーセンテージとしては分からないのですが、今回、5 年間で 65 試験、1 年全体を合わせると 10% 弱ぐらいの感覚になると推測します。

山本部長 全体の中で 10% がプラセボ対照をやっていると。

稲垣氏 それぐらいはやっているのではないかと、10% 弱、10% を超えることは

ないだろうという感じですが、毎年の試験数を足して、新たに試験届けを出す回数から考えますと。

山本部長

多くはないと。

稲垣氏

多くはないです。

山本部長

多くはないということですね。

稲垣氏

ただ、会社によっては結構プラセボ対照試験を積極的にやっていますと
言ってきたところもありましたので、会社の考え方にもよるところはあ
るかもしれません。

山本部長

実際に医療機関で拒絶されたというのは、医療機関が拒絶する。医療機
関の責任医師が拒絶するときと、今の時点だと倫理委員会が駄目だとい
うときがありますので。

稲垣氏

そこは回答として出ておりませんが、多分その両方を含めて。それで、
医師が拒絶という場合も、どうせ倫理委員会が通らないからというところ
で 駄目だと言ったのも入っているかと思います。

長野本部長

円グラフで2つぐらい質問したいのですが、回答6で、今、先生が少し
コメントされました児童の11歳未満は非常に苦労したと。それは何かコ
メントがあったのですか。

稲垣氏

子どもさんでの同意取得については、苦労するでしょうね。親御さんとし
ては早く治したいという希望があって、プラセボは避けたいという話
の中で、今度はプラセボ試験をお願いしたいと話をして、悪化した時は
レスキューの措置等を含めてあるからという条件でも、苦労されたとい
う話は直接伺いました。

長野本部長

親としては認めたくないけれども、ということで説得されたということ
ですかね。

稲垣氏

はい。

長野本部長

もう1つ、回答9でいろいろ、対照試験を実施した際の工夫があるかと思
います。私は不勉強でよく分かっていないのは、4番目ぐらいに「非釣
合型」と、これはどういうことですか。

稲垣氏

プラセボ群と対照群の患者の比率は1対1ではなく、プラセボ群のほう
は極めて少なくして、一応比較はできるのですが、プラセボに当たる率
は少ないですという形でやっているところがあります。

荒川副部長

精神神経領域といっても、対症療法薬と疾患修飾薬という今期待され
ている領域とかで、また、プラセボといっても飲み薬か注射薬かという
場合とかで、多分ハードルが随分違うのだらうと思うのです。実際、疾
患修飾薬の場合は、極立った即応的な効果が見られないと、脱落もかな
りあり得るという中で、どうコントロールしていくかという大きな問題

のような気がするのです。そういったことは評価委員会とか、そういったところで議論にはならないのですか。

稲垣氏 今回の回答 7 で 12 か月より長期などというのは、結構長いプラセボ対照というのもありまして、そういったのが疾患修飾薬というような、回復させるといよりは、進行を遅らせるようなタイプの薬剤の試験かと想像されるのですが、こういった場合の脱落に関しては、どうやって防ぐかは、特にその部分だけを回答、調査というか、ご意見を聴いてはおりません。ただ、下のところの何かあったときにはレスキューにもっていくということかという形の中で、後は試験を続けていただいたということかと思っています。

荒川副部長 レスキュー薬は基本的には対症療法的なものが多いような気がするのですが、疾患は進行していく。でも、もともと疾患の進行が緩徐な場合に、長期な試験が必要になってくるので、現われたときには少し非可道的な不利益が発生してしまうのが常だと思うのです。ですので、その辺のところは倫理的な問題と絡めてどう制御するかという問題のような気がします。

松田委員 がんは難しいかと思っていたのですが、がんで経験が多いということはどう考えるのですか。

稲垣氏 既存薬に対する上乘せという形で試験を行っております。そして、がんの場合、既存薬といいましても、今使っている薬に対して、今度は耐性が出てきて、薬に対して反応が悪くなってきた状態で、次の薬が欲しいという患者がおりまして、その場合、それを 1 群継続で、では、これはどうですかを上乘せでやるという形で試験を組みますので、そのためにどうしても上乘せでの効果としての試験が多くなります。

山本部長 こういうデータは非常に貴重なデータですが、国民性というか、日本の状況も 10 年前に比べると大分理解が進んできたと思うのですが、欧米、そのほかの地域と比べて今回のデータはどのような感じなのでしょう。

稲垣氏 特に最近の欧米のデータ等で比較はしておりませんので、申し訳ありません。そこは分かりません。想像としては、向こうのほうがやはりプラセボ対照が多いのかと。

山本部長 10%ではなく、もう少し多いことですね。それを打破するのはなかなか難しいし、打破すべきかどうかは分かりませんが、昔よりは。

稲垣氏 いや、もっと少ないかと思っていました。

山本部長 意外と上がったのは。

稲垣氏 65 だと 10%ですが、これ以外のところでも、各社は多いところで制限していますので、もっとやっているところはかなりの数でやっている

ころもありますので、昔よりは増えてきたのかという印象を持っております。

山本部会長 PMDA の職員の方で、海外との情報やこの情報と今回の何かコメントはありますか。

山田新薬審査第二部長 具体的なデータは持ち合わせていないのですが、基本的には FDA はプラセボ対照試験を求める傾向があります。

山本部会長 EMA は。

山田新薬審査第二部長 EMA は非劣性も認めています。

長野本部長 今回の発表で質問があるのではないかと思います。

野村新薬審査第一部長 2点ほど質問させていただいてよろしいですか。

長野本部長 はい。

野村新薬審査第一部長 今回、大変貴重な調査をしていただきまして、ありがとうございました。1点目ですが、回答7に対して、8とか9、投与期間が、今話があったとおり、比較的短いものからそれなりに長いものもあったと思うのですが、比較的短いものと長いもの間でそれぞれ、例えば試験のデザインに傾向があるのかとか、あるいは工夫の仕方、先ほどもレスキュー薬のお話とか、途中で悪化をした方々が中止をするものがあったと思うのですが、試験期間との相関みたいなクロス集計みたいなことは。

稲垣氏 そのクロスの中では、何か一定の傾向みたいなものは特に見られませんでした。ちょっと例数が少ないところもあると思います。

野村新薬審査第一部長 もう1点、これもお分かりになればということになりますが、7ページの上のほうで「実施段階での問題点」で、IRB からのご指摘により「承認保留」ということが記載されているのですが、一般的に IRB がこういった試験をよろしくないとおっしゃるような場合ということとは、どういう場合があるか、もしもアンケートなりからお分かりになることがあれば。

稲垣氏 アンケートでは特に出ておりませんが、先ほど少しお話がありましたように、ある程度長い試験の場合、進行についてどう考えるのだということとは、いろいろと話はあるようです。

田代委員 これもアンケートの結果ではないと思うのですが、IRB からの指摘事項で確かに悪化のリスクの話は必ず出るとは思うのですが、それ以外にご本人に同意能力があるか、ないかということによって指摘が出てくる場合は、あるのかどうかということですが。

稲垣氏 小児の場合。

田代委員 小児ともう1つ、精神疾患の場合で特に同意能力がない方も、プラセボ試験が入ることに対して拒否反応があるのか、ないのかということでは、

聞いたお話もあるのですが、そういうことも一応指摘事項としてはあるのですか。

稲垣氏 今回の回答として直接はありませんでしたが、IRB はいろいろな観点から同意取得、あるいは何かレスキューの話、それを出そうかどうかとして、進行をどうやってチェックするかというところまでも、測定の進捗とか、いろいろなところでご指摘をいただいているという話だけは聞いております。

山本部会長 私が関与した治験の中で、全国で30施設ぐらいのIRBにかけて、2施設は絶対駄目だと言われて、そこだけ脱落したというのがある。我々の施設もそうでしたが。全然判断が違ってきてしまうというのですね。そういうことがあります。だから、そういう意味では非常に難しい問題で、イエスかノーの問題ではないということはよく分かるのですが。

よろしいですか。もし、よろしければ、とりあえず次の佐藤先生のプレゼンをいただいてから、また総合討論がありますので、どうぞよろしくをお願いします。次の議題として、「医療統計の視点からのプラセボ対照試験について」ということで、前回までの議論で実薬と非劣性試験における問題点とか、プラセボ対照試験の設置の意義について、医療統計の先生から1度話を聴きたいということがありましたので、委員の中にいらっしゃる佐藤先生からお話を聴くことになりました。佐藤先生、どうぞよろしくをお願いします。

< 議題2：医療統計の視点からのプラセボ対照試験について >

佐藤委員 京都大学医療統計の佐藤です。第2回プラセボ専門部会のときに、浜松医科大学の渡邊先生からこのスライドで非劣性試験、同等性試験を続けていくと、プラセボと有効性に差がないものが出てくる可能性があることをお示しされて、これはいったいどういうメカニズムで起こるのかを説明してほしいというお話でしたので、今日、部会長からは学部レベルの講義をというお話だったのですが、さすがに学部で非劣性の試験は無理なので、大学院レベルの話をさせていただきたいと思います。

初めに、そもそも「非劣性試験」という言葉がどこで出てきたのかというと、この部会でもICH E10、対照群の選定のガイドラインが前に紹介されていましたが、最初に出てきたのがICH E9の統計ガイドラインでこの用語が定められました。そのときに優越性試験、同等性試験、非劣性試験という3つに分けようということが決まりました。最初に簡単にそのお話と、今の非劣性試験をつないでいくと、プラセボと同じものが出てしまうことは古くから知られていて、「バイオクリープ」と呼ばれて

いるのですが、バイオクリープという現象がどうして起こるのかということ。それを解決するのはそれほど難しくはない気がするのですが、ところが、非劣性試験特有の問題点があって、なかなかうまくいかないのだということ、まとめてお話ししたいと思います。

最初に、ICH E9 の統計ガイドラインですが、それまでは普通の試験、優越性を示す、有効性を示す試験はよくやられていて、今でいう非劣性試験は同等性試験という名前で、同等性と非劣性が一緒くたに扱われていました。それをきちんと分けたほうがいいだろうということで、E9 の統計ガイドラインの中で Superiority Trials、Equivalence Trials、Non-inferiority Trials ということで試験のタイプを 3 つに分けることを確認しました。

優越性に関しては、皆さんはもちろんよくご存じだと思いますが、普通の検定を行います。帰無仮説としては、新薬はプラセボ又は実対照薬と等しいという仮説に対して、新薬候補は有効性で優れているのだということを検証する試験です。

統計的に有意な結果になれば、新薬候補が勝っているのだと判断して申請してくるわけですが、有意でなかったときに、では有効性は同等かということが問題になります。このことが問題になるのは、検定が有意になるということに、2 つ意味があります。1 つは、新薬候補が本当に対照薬あるいはプラセボよりも優れていて、統計的にも有意になるという場合、これは問題ないのですが、もう 1 つは、サンプルサイズが非常に大きくなったがゆえに、ほんの小さな差を検出してしまって、統計的には有意になる。でも、臨床的にはほとんど意味がないという場合があります。

ということは、裏を返すと、サンプルサイズが非常に小さい試験を設計すると、統計的に絶対に有意にならないわけです。このときに同等だと判断してしまうと、実は新薬候補が対照薬より劣っている場合であっても、統計的にサンプルサイズが小さい試験を行うと、統計的には有意になりませんから、この場合でも同等だと言えることになってしまいます。ですから、検定で有意でなかったから同等だというロジックは、明らかに間違いです。ただ、残念ながら日本でも 1992 年に厚生省が統計解析ガイドラインを出しているのですが、このガイドラインが出るまでは統計的に有意でないから同等だという論理がまかり通っていて、新医薬品が承認されてきました。これが 1992 年の統計解析ガイドラインの施行によって、このことは一応プロテクトされてきたわけです。

では、有意ではないから同等だと言えないのだとすると、どうしたら

等しいということを経験的に証明できるかが考えられてきたのですが、1つの考え方としては、等しいわけですから、信頼区間を計算すると、それは0を含まないとおかしいですね、そうでないと等しいことになりませんから。信頼区間を計算するのですが、それが0を含んでも、青いバーのように幅が非常に広いと、どこにあるか分かりませんから、これはとても等しいと言えないのですが、もし、それが右の赤い試験のように非常に狭い範囲に収まってくれば、これはほとんど同等と考えてもいいのではないかと分かります。つまり、信頼区間を計算して、その幅が0を含んでいて、非常に狭い範囲に収まっていれば、有効性はほとんど2つの薬で同じだと考えてもいいのだらうということになります。

ただ、問題は、では、どのくらい狭かったら同等だと判断していいのかということが問題になります。この同等であることを証明するための試験、先ほどのICH E9の用語で言うと同等性試験になりますが、同等性試験では、この範囲に入ったら信頼区間がこの範囲に収まっていれば、2つの薬は同等だとみなしてもいいという範囲を事前に決めます。これを「同等マージン」と呼んでいるのですが、マージン幅を取って、この幅に信頼区間が収まったら、2つの薬は有効性が同等だと判定しましょうと決めます。例えば、有効割合の差が $\pm 10\%$ とか、 $\pm 7\%$ とか、何らかの決め事、観点から決めなければいけないのですが、そういったものを決めます。これは単なる決め事ですから、それで本当に臨床的に同等かどうかはまた別の話ですが、いずれにしてもこういう範囲を決めないと、同等であるかどうかは判断できないことになります。

優越性試験と同等性試験、今の話について信頼区間を使って図に模式的に表してみますと、0%より上のほうに行ったら新薬が有効で、下のほうに行ったら対照薬が有効ということにしますと、例えばある試験をやって、信頼区間を計算して、このような信頼区間の幅になれば、これは信頼区間の下限が0よりも上で新薬有効のほうに行っていますから、これは優越性が示せた試験と判定することになります。

それに対して0を含んでいる信頼区間、あるいは、一番右の試験のように完全に0よりも下回ってしまっている試験、こういったものは統計的には有意ではないという結果になります。それで、一番右の試験だけが優越性を示すことができるということになるわけですが、では、同等性試験の場合はどう考えるかといいますと、先ほどお話ししたように、同等性マージンを決めますから、例えば、先ほどの例のように $\pm 7\%$ の範囲に収まっていれば、2つの薬は同等だと判定することにすると、 $\pm 7\%$

のところは線を引いてやればいいわけです。同等マージンの範囲の中に入っている試験は、同等性が示されたこととなります。この例でいきますと、真ん中の試験は、今、赤にあります、すっぽり同等マージンの中に入っていますから、この試験だけが同等性が示せたこととなります。後の試験は、先ほど優越性が示されたと言った試験も、これは同等マージンの中からはみ出ていますので、ほかの 4 つの試験の結果は、全て同等性は検証できなかったこととなります。

積極的に同等であることを示したい試験もありまして、それは例えばゾロ薬です。ジェネリックとか、あるいは遺伝子組換え製剤などの場合は、これは完全にコピーをしているわけですから、同じものを作っているわけですから、元の薬よりも良くても悪くてもいけないわけですので、これはきちんと同等性を検証しなければいけないわけです。

ただ、普通の薬でこういったコピー薬でない場合には、劣りさえしなければ、良い方向に出る場合には、それは認めてもいいのではないかと。これは自然な要求だと思うのですが、劣りさえしなければ、良い方向に行く分には、それはウエルカムだろうということで、そこで同等性だけではなくて非劣性試験、一定以上劣らないことを担保する試験という考え方が出てきました。

非劣性試験ですが、今度は同等性試験と違って、同等マージンのように±何とかというふうには上と下で押える必要はなくて、下限だけを決めればいいわけです。例えば非劣性マージンと呼ばれるものを下側の -7%のところだけに設定することにします。そうすると、このような形になって非劣性マージンが -7% になりますから、信頼区間がこれよりも上に行っている試験は非劣性が証明されたこととなります。

今、一番右の緑の試験だけ「優越性」と書いてありますので、これも非劣性試験の範疇から言うと非劣性が証明されたこととなります。あと、先ほどの真ん中の赤にした同等性が示された試験と 1 つ右、今、赤になりましたが、この 2 つの試験は非劣性マージンよりも信頼区間の下限が上に行っていますので、この 2 つの試験は非劣性が検証できたと考えることができます。これが非劣性試験の考え方になります。

非劣性試験のマージンの考え方としては、同等性でもほぼ同じですが、実対照薬よりも一定以上劣ったら、それは薬として有効性として許容できない限界であるということから、臨床的な観点から決められますが、1 つの考え方としては、最初に実対照薬が承認されるときには、多分プラセボ対照試験を行って承認されているわけですので、少なくとも実対照薬とプラセボとの有効割合の差よりも小さくしないと、マージンがもし

プラセボよりも超えて設定してしまうと、バイオクリープどころか、いきなり有効性がプラセボよりも悪いものが承認されてしまう可能性もありますので、少なくともそれよりは狭くすることが通例になっています。例えば、プラセボとの差の有効性の 2 分の 1 の幅をマージンとしますとか、あるいは 3 分の 1 をマージンとしますというふうに、必ずプラセボとの差よりも狭い範囲を非劣性マージンとして決めることになります。

当然ですが、マージンを小さく取ると、必要なサンプルサイズ、参加者の数が多くなりますし、マージンを大きくすれば、サンプルサイズが小さく、対照者数を少なくすることができますので、こういうことを言っただけですが、企業としては、できるだけマージンを大きく取ってサンプルサイズは小さくしたいのでしょうし、規制当局としては、できるだけマージンは小さく取って厳密に試験をやってもらいたいというせめぎ合いになるかと思えます。

非劣性試験のもう 1 つの例ですが、今、これは明らかに、有効性が対照薬よりも新薬のほうが劣っている試験があり、一番右端の 0% よりも信頼区間が下に行っている試験ですが、これも信頼区間の上限が 0% より下に行っていますから、新薬のほうが有効性で対照薬よりも明らかに劣っています。ただ、今度は、先ほどの設定した -7% という非劣性マージンよりも信頼区間の下限が上に行っています。ですから、この試験は有効性では統計的に劣っているのですが、許容した非劣性マージンの範囲内では、この試験でも非劣性が検証できたと判断することになります。これが非劣性試験のいやらしいところで、明らかに有効性で対照薬よりも劣っている新薬が、非劣性試験のマージンの設定によっては、非劣性で承認されることも起こり得るわけです。

これが実例ですが、2013 年に承認されたアクテムラ皮下注という薬です。これは第二部会で山本先生もいらっしゃって承認されたのですが、関節リウマチの抗体薬です。もともと点滴静注の製剤が承認されていたのですが、皮下注ですと自宅でできますので、利便性に勝るということで、点滴静注製剤との非劣性試験が実施されました。最初に点滴静注製剤が承認されたときの試験の成績が、24 週の関節リウマチの改善率の差が、標準的に使われているメトトレキサートとの差が 57.3% と、非常に著効している薬ですが、プラセボとの差が、これは 12 週での差ですが 65.5% というところで、圧倒的に勝っていました。

そこで、例えば先ほどのプラセボとの差の 2 分の 1 とか、3 分の 1 とかをマージンに設定するという話がありましたが、この試験ではメトトレキサートとの差の 3 分の 1 が 19.1% だったので、非劣性マージンとして-

18%を非劣性の限界として設定しますという試験が行われました。実際にこの試験の結果がどうだったかというと、24週の改善率が皮下注群で約80%、点滴静注群で88.5%ということで、改善率の差が-9.4%、95%信頼区間が-17.6~-1.2%ということで、右のボックスの赤い線で引いてあるところが非劣性のマージンの限界のところですが、今の信頼区間と、点推定値をプロットしますとこのような形になって、非劣性マージンが-18%ですから、信頼下限が-17.6%で、ぎりぎり非劣性が言えているのですが、上限も0を下回っていますから、これは明らかに有効性で点滴静注製剤よりも劣っている薬だということになります。

この薬に関しては利便性のほうが優先するというので、点滴静注製剤に比べて10%ぐらい改善率が劣る可能性はあるけれども、そのことを十分周知した上で利便性に勝るので承認しようということになりました。ただ、機構としても、非劣性マージンの-18%はプラセボよりは上に行っていることは担保できているけれども、点滴静注製剤と皮下注製剤との非劣性という観点からは、過大に広く取り過ぎたのではないかという懸念はあることを、審査報告書にまとめられていました。少し劣っている薬は、非劣性マージンの設定の仕方によっては承認されてしまうことが十分起こり得るわけで、実際にこういう例もありました。

マージンの範囲内で劣った新薬候補が承認されてしまうことは、これは非劣性試験の性質上しょうがないのと、先ほどお話ししたように、サンプルサイズは非常に大きくすれば、どのような小さな差でも有意にすることができますので、非劣性試験でもほんの少し負けている薬であれば、サンプルサイズは非常に大きくしてやれば、非劣性マージンよりも上回るサンプルサイズを設計することができて、劣った薬でも承認されてしまうことになるわけです。

そうすると、少し劣った医薬品が次の実対照薬になって、また新しい新薬候補と非劣性試験を行う。また、それが少し劣ったものと新しい新薬候補で非劣性試験を行う。こうやって順々につないでいきますと、最初の渡邊先生が使われたスライドにありますように、最終的にはプラセボと変わらない医薬品が承認されてしまうことも、原理的には起こり得ると思います。

ただ、これが起こり得そうもないということは当然ですが、先ほどのアクテムラの皮下注の例でも非劣性マージンが-18%と、我々とするとは異常にと言っているぐらい大きい幅ですが、そのように大きい幅に取らなければ、もっと厳しい幅にしておけば、例えば先ほどの例でも、-15%ぐらいにしておけば非劣性は検証できなかったという結果になりますから、

マージンを厳しくすれば、こういったプラセボと同等なものが承認されてしまうことはまず起こり得ないはずですが。

ただ、そうはいても、非劣性試験に特有のいくつかの問題点があって、これを解決することがなかなか難しいとされています。1つはICH E10、対照群の選定のところのガイドラインに詳しく書かれている分析感度という問題で、分析感度は実対照薬がプラセボに必ず勝っているのだという保証があるということ。これはプラセボを試験の中に組み込んで、先ほど稲垣さんのお話の中でも、プラセボと実薬と新薬候補と3剤併用という話がありましたが、プラセボが入っていれば必ず証明できるのですが、プラセボがないと、もし新薬の候補も実対照薬もどちらも効かない患者を非常にうまく臨床試験に選んできて、その患者ばかり組み込んだとすると、どちらも効きませんから、ほぼ同じ成績になって、どちらも効いていないのに同じだということが証明できてしまうこととなります。プラセボが入っていれば明らかにプラセボとも同じ結果になりますから、それはおかしいことが分かるのですが、プラセボが入っていないと、こういうことが言えてしまう。

もう1つの問題は定常性と呼ばれている問題で、最初に実対照薬が行われたときのプラセボ対照試験での成績は、その後引き続き行う非劣性試験のアンカーになっているわけです。ところが、プラセボ対照試験で検出された有効性は、かなり昔の試験になりますから、長い年月がたっていて、支持療法とか、いろいろな背景の医療状況が変わって、当時の、あるいは10年以上前のプラセボ対照試験の結果は、本当に現在でもその有効性が再現できるのか、それが再現できるのが定常性となります。この定常性が示されないと非劣性試験は担保できませんので、いろいろな支持療法とか、治療技術の進歩でそういったことが言えなくなってしまうと、もともとの非劣性試験自体のアンカーになっているプラセボ対照試験が崩れてしまうこととなります。

まだいくつかあるのですが、最初に非劣性マージン自体も、プラセボとの差の2分の1とか、3分の1とかというところでマージンが決められますから、最初に実施したプラセボ対照試験は非常に重要な結果ですが、ところが、プラセボ対照試験のデメリットとして、まずそれと同じデザインで実施する必要があるのですが、先ほど言いましたように過去の試験ですので、いろいろなことがあって、過去の試験と同じデザインをずっと繰り返して何年も実施するのは難しいことがあります。

もう1つの問題は、プラセボ対照試験で統計的に有意になるためには、サンプルサイズがそれほど大きくななくても、プラセボとの差ですから、

そこは大きな差がつきますので、サンプルサイズが小さくても検証できます。そうすると、今度は治療効果の推定という意味では、サンプルサイズが小さいことによって、効果の推定の精度が非常に不安定になります。信頼区間が広がってしまいます。そうすると、本当に最初に検証したプラセボ対照試験の有効性の差は、また再現可能かという問題が起きてきます。

もう 1 つは、時間が経つことによって、主要エンドポイントが、医療技術とかのいろいろなものの進歩によって、エンドポイント自体が変わってしまうことが起こります。そうすると、最初に実施したプラセボ対照試験と同じデザインで実施する必要があるにもかかわらず、もはや時代の流れとともに同じエンドポイントは組めなくなってしまうという問題が起きてきます。

最後ですが、こういった臨床試験には付きものですが、ノンコンプライアンスです。割り付けられた薬をきちんと守らないとか、試験の途中で脱落してしまうということの影響が、こういったプラセボ対照試験でも、普通の優越性試験よりも同等性試験や非劣性試験といった試験に与える影響が大きいと言われています。ICHの統計ガイドラインではどう書かれているかと言いますと、同等性試験又は非劣性試験は、本質的に保守的でないのです。試験の計画上又は実施上の多くの不備が、先ほどお話ししたノンコンプライアンスとか脱落の問題ですが、それが偏りを生むのですが、基本的には同等であるという方向に臨床試験の偏りが生まれます。このことは優越性試験では優越性が示せないという方向に行きますので、大きな問題にはならないのですが、同等性試験、非劣性試験では、同等であるという方向にバイアスがかかってしまうと、劣っている薬でも同等であるというふうに検証ができてしまうことになりますから、大きな問題になります。ですから、登録基準違反、服薬不遵守、治験治療の中止、追跡不能、欠測データ、治験実施計画書からのその他の逸脱を、こういった非劣性試験あるいは同等性試験を実施する場合には、最小限に抑えることは特に重要であり、また、これらが解析に与える影響も抑えることが重要であることが、統計ガイドラインにまとめられています。こういったことを受けて、E10の対照群の選定では、様々な試験デザインが提案されていて、非劣性試験だけではなくて、先ほどお話しにもあったプラセボを入れた 3 剤の比較試験とか、早期のエスケープ試験とか、様々なデザインが提案されているわけで、プラセボ対照試験が一番キーになっている、アンカーになっているということで、その成績が再現できない、あるいは担保できない状況になったときには、非劣性試験

だけではうまくいかない場合があることが、往々にしてあることになります。以上で私の話を終わります。

山本部長 佐藤先生、ありがとうございます。数式が出てこなかったのが、よく分かりました。最初のときのバイオクリープですか、特に精神神経疾患のところでも良いと言われているものも含めて、何となくイメージとして非常によく分かったという感じがしますが、先生方、ご意見ありませんでしょうか。

佐田委員 非劣性マージンの決め方によっても、いくらでも変わるということで、そのあたりが実薬との2分の1とか、先ほどのアクテムラときには3分の1とか、その辺はどうやって決めるのですか。

佐藤委員 これはあくまでも決め事なので、もう本当に、規制当局とのネゴシエーションだと思います。基本的には臨床的に決められれば一番良いのですけれども、この薬だったら今の場合にはプラセボとの差から-3分の1の-18%と決めましたけれども、実際には本当にこれが非劣性を証明するためには、やはり-10%が限界だろうとか、-7%が限界だろうということが臨床的に議論できるのが一番良いのだと思います。ただ、なかなかそれもコンセンサスを得るのが難しいので、機械的にと言いますか、何らかの規準を決めざるを得ないというのが、現状だと思います。

山本部長 先ほど佐藤先生はせめぎ合いだとおっしゃいましたが、実際にやはり、企業側とPMDA側はせめぎ合っているのですか。

山田新薬審査第二部長 まさに、そのとおりだと思います。実際、どうやって決めるかというルールがありませんので、基本的には2分の1にするか、3分の1にするか。その辺に落ち着きます。

佐藤委員 企業もいけると思えば厳しくしますし。

山本部長 そういうところのガイドが、もしこの科学委員会で少しでも言えたらいいなという気がしますが、我々もなかなか難しいところがありますね。先ほどの1つの例では、アクテムラの例ですけれども、プラセボとの差を証明するための試験ではなくて、今まで点滴静注が入っていて、それと代わるかどうかということなので、あれをプラセボよりもいけますよね、というのだったら、それはやはり正しくないことになりますよね。そのところの設定は少し大き過ぎたかなという気がしますけれども、稲垣先生どうですか。

稲垣氏 今の、まさにせめぎ合いのところ、我々の業界の中でもよく話が出るのが、佐藤先生の資料の中でもありましたが、支持療法がどんどん変わってきたことで、プラセボとの差というのが最近のデータが必ずしも得られないような場合もあって、その後の環境の変化で昔の数字を基準に

してもいいのか、それともどうなのかというところが、中で聞いていると、結構きわどい話をするというようなところになるのかなと、中で聞いているところです。このあたりが、考え方、あるいは新しいデータが常に得られているところであれば、しやすいのですけれども、それがなかなか得られないと、どこが正しい値なのか、どちらも決めかねるという感じに、実際そうなるので、せめぎ合いという話になってしまい、まさにそのとおりなのかなと、そんな感想を持っております。

荒川副部長 今の話の続きになるかもしれませんが、時代とともにというばかりではなく、国際共同試験の中では国によって医療環境も異なっていて、そのことも実際に大きな要因です。一昔前にですけれども、胃癌を対象としたアバスチンのアバガストスタディのときには地域差を見ていて、アジアでおそらく日本、韓国その辺のところだと思いますが、そのあたりでは全生存期間で差が付かない。

ところが、パンアメリカンと言っているような、おそらく南米といったところだと思うのですが、そういったところでは有意差がつく。何か違うかということ、コントロールアームが違うのですね。コントロールアームがパンアメリカでは短い。その地域だけで有意差がついてしまう。ところが全体では付かない。これはおそらく医療環境の差によるもので、試験に入る段階で、もう既に患者さんの質も違うでしょうし、それからプログレッション（がんの増悪）になった後の治療も違うでしょうから、そこはどのように解析するか難しい問題です。今後そういった医療環境の差ということも、デザインの上では十分考えていかななくてはいけない問題だとは思いますが。

山本部長 試験自体は正しくて、要するに、内的には妥当だけれども、それが外的妥当性を持つかどうかというのは、この領域にはどうしようもないことなのですか。それを何とか、全体にやる学問はあるのでしょうか。

荒川副部長 私に聞かれても困りますが、要は、ちゃんと統計学的に何らかの形で有意であると。ある条件の下でということになってしまうかもしれませんが、それを出していくことが1つですが、先ほど佐藤先生からもご指摘があったように、サンプル数を大きくすれば、当然、有意差が付く可能性が高くなるのですけれども、本当に臨床的に意味があるのですかということは、今後考えていかなければいけなくて、5,000例、1万例のスタディをやらなければ差が付かないようなものに対して、NNTのような医療経済的な考え方も当然入れるべきで、そういったスタディを受け入れるかどうかという問題を当然入れるべきで、今後はそれについての議論があってもいいと思います。

ですから、人の命にはなかなか変えられないというところで、無制限的に考えるのか、それとも、1万例もやらなくてはいけないようなスタディで有意差を出すのは、少なくとも主要エンドポイントでは、あまりこれからは受け入れられないと考えていくのか。そういった実質的なサンプルサイズの面からも議論があってもいいと思います。

山本部長 骨粗鬆症などでは、圧迫骨折の数に入るのでは。

荒川副部長 そうですね。前回の議論の中では、どうしてもハイリスクの患者さんを入れざるを得ない。アルツハイマー病でもそうですし、皆、発症直前のところ、あるいはハイリスクで、非常にイベントの率の高い方を選ぶと。そういうスタディも重要だとは思いますが。

けれども、その問題はもう1つは、本当は薬剤の性質からすると、もっと早い段階から治療をしなくてはいけないのに、実際に承認を得られるのはハイリスクの患者さんだけという実態があって、これをどうするかということは今後の一般化可能性の中で議論していく問題だと思います。それはもう1つの問題だと思います。

山本部長 その辺までずっと議論してしまうと、ちょっと行き過ぎてしまって、我々の今日の議論のポイントから外れてしまうと思いますが、でも、明確な結論はなかなか出ないということは、まず知っておきながら、なるべくプラセボのところの特化して議論したいと思いますが、まず、佐藤先生のプレゼンテーションに何か先生方どうでしょうか。

荒川副部長 やはり、佐藤先生が言われたマージンをどうするかというのが、イメージとしてもっとはっきりと皆さんでも分かって、科学的なものがあるといいなという感じはしますよね。少なくとも、その根拠をきちんと書いておいていただければいいのかもしれないけれども、難しいところですね。

稲垣氏 根拠とできるデータが得られるかどうかというところが、結構大きいのかなと思っておりまして、プラセボ対照試験がいつ実施されたのか。そして、主要エンドポイントの取り方等も時代で変わってきている中で、今回新たにやろうとしている試験と参考になるものがあるかどうかで、何%という数字、もともとのプラセボとの差がどれくらいかという計算ができるかどうかというところが大きいのかなと。そういうときに、やはり企業として、本来プラセボ対照試験をやったほうがいいのかもれないと、ずっと悩むところではあるのですけれども。

< 議題3：その他 >

山本部長 一応、佐藤先生に対してのディスカッションの後は、2つの稲垣先生の

プレゼンも含めて、今日の全体のことで少し議論をしてみたいと思っていますが、両方を含めての話題で結構ですので、何か問題点をお願いしたいと思います。

それから今後のことになりますが、そろそろ第4回が終わりますので、次からは、まとめの方向になりますので、その前に、これが足りないというようなことが、もしあれば、それはご指摘いただいて、そうでなければ次回からはまとめの方向ということになりますので、第1回に何をやったのかというのは、なかなかすぐに頭に浮かんでできませんが、また議事録が必要であれば、次回、議事録を皆さんに見ていただきながらということになると思います。

プラセボ対照試験はどうでしょう。今日は問題点が分かったというのが、まず分かりましたね。マージンの取り方だというのは大きな問題であるということは分かりましたけれども、それをどう解決するかというのは、なかなか難しいということも分かりました。

何にせよ、どこか方向性で何かというのは、ありますか。もうちょっと我々、議論したほうが良いという。

長野本部長 今まで4回議論してきて、プラセボ対照試験をまず基本とするというところはあるかと思っています。また、実際にその試験の中身に関して、今の佐藤先生のお話にもありましたが、取り方によっては、いろいろ問題点も出てくる。随分、実際には同等性試験だけではなくて非劣性ということで、以前も指摘されて、今回そのポイントを十分に把握できたわけですが、その中の試験に非常に問題があるといった、大きく分けてその2つの方法があるのだろうなという感じがいたします。

実際にそういった対照試験が取れないという例も小児から始まって、超希小疾患等も随分、指摘されてきましたので、その辺に対して実際に企業の現場ではどうなっているのかというのは、今回の統計の円グラフの中でも、小児は非常に極めて少ないということですよ。それは十分に予想できたわけですが、その辺の問題点が、かなり明らかになってきたのかなという感じはいたします。

山本部長 稲垣先生のおっしゃられたように、イメージとして全体の治験の約10%程度がプラセボだとすると、90%がそうではない非劣性か同等試験をやっているわけですよ。ということは、マージンを決めるために毎回せめぎ合いをしているということですか。

山田新薬審査第二部長 全体の治験ですと、非盲検非対照試験もあります。

山本部長 非盲検試験というと、イメージとしてはどんな形になりますか。上乘せだけですか。

山田新薬審査第二部長 いや、比較せずに、そのまま。

山本部会長 比較せずにそのまま。ワンアームか。

佐藤委員 コントロールなしで。

山本部会長 コントロールなし。

佐藤委員 コントロールなしで比較せず、安全性確認試験みたいなものは、ワンアームで得られることであろうかと。

山本部会長 それはどのように後で統計的な処理をするのですか。

山田新薬審査第二部長 基本的には集計だけで、今おっしゃった長期安全性試験のような。

山本部会長 ああ、長期試験のような。

佐藤委員 安全性を確認するのが目的です。

山田新薬審査第二部長 相試験や 相試験も含まれます。

山本部会長 なるほど。それはちょっと今回の議論の対象外ということになりますよね。それが大きな割合を占めているのですか。

稲垣氏 かなりの時間はございます。長期試験は必ず実施しますので。

山本部会長 そうですね。普通に治験と言われている P1 から P3 までの間で。

稲垣氏 長期試験も治験として実施しております。

山本部会長 その中で、このマージンを決めなくてはいけないという問題が、どのぐらいの割合で生じているのでしょうか。

野村新薬審査第一部長 補足になりますが、先ほどのマージンのお話は、毎回せめぎ合いなのかというご指摘がございました。おそらく、領域ごとに過去に既に薬が承認されているような分野においては、審査報告書なども公表しておりますので、おおむね各社は過去の品目が、マージンをどのように取ったのか、というあたりも横目で見ながら設定をしていくことが多々ありますので、ただ、それが本当に何か意味のあるというものではないのですが、その疾患において過去に使われたことのある数字ということと、そういったものを踏襲しているようなケースはよく見られるかと思えます。

山本部会長 それは実際にはよく我々には分かるのですけれども、科学委員会として、国民にも分かるマージンの引き方というのがないと良いなという気がするのですけれどもね。

それが段々、バイオクリープの方向に行ってしまうのが、良くないですよ。ここをどのようにするかというのを、しかし、マージンの設定だけですものね。この考え方はもうこれ以上、我々が数学者や統計学者ではない限り、新しい方法を見つけるわけにはいかないのです。これはここでは、どうマージン設定するかということだと思いますが、佐藤先生、

何かそこに対するご意見は。

佐藤委員 なかなか、これ、先ほどちょっとお話ししましたけれども、1992年に統計解析ガイドラインというものができたのですけれども、その注のところに目安として10%と書かれていたのですね。ですけれども、その前にあくまでも目安で、それは各薬効領域とか臨床領域できちんと決めましょうということになっていたはずなのですけれども、それがそのままずっと何もされないまま来てしまったというのがあって、必ずしも、多分、製薬メーカーだけの問題ではなくて、やはり学界などできちんと議論していただくべきものなのだと思います。

私は基本的には臨床的に決めるのが一番いいのではないかと思っているのですけれども、逆に臨床の先生たちは統計的に決められないのかということをおっしゃっていて、これもずっとせめぎ合いが続いています。

山本部会長 臨床的に決めるというのは、おそらくサンプルサイズがどれくらい可能かということにも、かなりよってくるので、たくさん患者さんがリクルートできるというような状況だったら、それはやはり、たくさん入れましょう。そうすると、マージン低くても大丈夫だということはあると思うのですけれど。これはこれで置いておいて、そのほかはよろしいですか。

できたら少し、外国との比較などの情報が今日以上にあると、全体としての議論としてはいいのかなという気がしますが。

長野本部長 どなたかいますかね。

山本部会長 難しいですね。先ほど言われたようにFDAとEMAでも、もう姿勢が全然違うと。FDAはなるべくプラセボ対照を求めているけれども、EMAは非劣性を認めていると。そのスタンスの中の根拠はどのように議論をした上にこうなっているのかというのが少し分かりますと、ではPMDAはどうするかという方向が1つになるかもしれませんが、少し、その辺のデータはありますか。

山田新薬審査第二部長 類推ですけれども、FDAでは、とにかく、その新薬が要するにプラセボに勝っていれば、薬として認めるという方針です。

山本部会長 それは勝っていれば認めるわけですね。

山田新薬審査第二部長 EMAではどちらかと言うと、既承認の類薬と比較してどうかということが議論になりますので、非劣性試験を求めているのかなという気はします。

山本部会長 そうすると、同じ薬をアメリカで治験をやったときと、ヨーロッパでやったときとで承認の度合いが違ってくると。

山田新薬審査第二部長 ただ最近はほとんど国際共同治験でやられていますので、ほと

んどがプラセボ対照試験に移っています。それで、国際共同治験で非劣性試験というのは、実は、あまり見たことがないです。

山本部会長　　そうですか。それではよろしければ、こんなところで今日の議論を、ここまでにしたいと思います。それでは事務局から連絡事項をよろしくお願ひします。

蛭田事務局長　当専門部会につきましては、次回以降も基本的には、いずれかの月の第2金曜日に開催させていただきたいと考えております。詳細につきましては、後ほど事務局からご連絡を差し上げたいと思います。よろしくお願ひいたします。以上でございます。

< 閉会 >

山本部会長　　おそらく、来月はないだろうということで、できましたら次回からまとめの方向に。ですから、次回までの間に小さなワーキンググループでたたき台を作って、それをまず見ていただきながら、先生方に議論していただいて、またワーキンググループで少し文章化していく作業になりますので、少し、実際の専門部会の間隔は長くなるかもしれません。

このような流れで、よろしいでしょうか。もしよろしければ、その方向でこれ以降を進めさせていただきます。それでは本日の専門部会はここまでとさせていただきます。皆さん、どうもありがとうございました。