

ター(CRC), 看護師などの“local user”のほかにも, データセンターや運営事務局の中央支援機構の統計家やデータマネージャー, モニターといった“central user”, さらに, プロトコル審査委員会や効果・安全性評価委員会, 施設の倫理審査委員会の委員といった“reviewer”としてのユーザーが想定される。プロトコルとは, このように多職種かつさまざまな役割の人間が臨床試験を遂行していく上で用いる「取り扱い説明書(instruction manual)」であり, これらのすべての人に理解可能で, さらに, それらの人たちのニーズに応えたものでなければならない。プロトコルは該当する疾患の専門家が作成することが多いため, とかく難解で専門用語が羅列されているという印象があるかもしれないが, 専門用語を用いるのなら, 理解を助ける注釈がなければならないし, 内容が論理的で読みやすくなければ, ユーザーフレンドリーとは言えず, コミュニケーションツールとして機能しない。

とくに, この“reviewer”としてのユーザーには, 患者団体の代表やあるいは患者自身がなり得ることを念頭に置く必要がある。もちろん患者が読み手の中心ではないため, 過度に非医療従事者向けの表現にする必要はないが, 少なくとも患者が読んで不快に感じるような表現は避けるべきである。JCOG試験においては, 求めがあればプロトコルを試験に参加する患者に渡すことも許容しているため, 患者が不快に感じる可能性を最小化すべく, たとえば「症例」はできるだけ用いず, 文脈に応じて「患者」や「の場合には」など, 他の表現に置き換える工夫を行っている。

2. 品質管理ツール

PiantadosiのClinical Trials²⁾には, プロトコルとは「臨床試験を行うにあたってもっとも重要な品質管理ツールである」と書かれている。多施設共同試験では, 異なる施設であっても, 同じ適格規準を満たした患者に対して, 同じ治療, 同じ治療変更規準, 同じ評価が行われることが原則である。施設によってばらばらの適格規準によりさまざまな背景の患者が登録され, 施設独自のスケジュールや投与量で薬剤の投与が行われ, 担当医ごとの治療変更規準により勝手に減量あるいは休止が行われた場合, その臨床試

験の結果, たとえ「有効であり安全である」と結論されたとしても, 「どういう治療が」有効であり安全であるのかが記述できない。つまり, その「有効性と安全性」を日常診療で再現することが不可能であることから, 患者に役立つエビデンスを創ったことにはならない。われわれはこうした試験を「Study for study」と呼ぶ。いくら仮説が正しかろうと, 臨床試験の「質」が悪ければ, 正しい結果や意味のある結論は導かれないのである。臨床試験の「質」とは, 結果・結論の「信頼性」であり, その社会的・科学的な「価値」である。

こうした臨床試験における「質」を保つためのさまざまな工夫・活動を, 臨床試験の「品質管理(quality control)」と呼ぶ。品質管理とは, 「臨床的仮説→プロトコル→治療→評価→解析→結論」と続く一連の臨床試験の流れのすべての段階で生じ得るエラーを最小化する作業のことであるが, 臨床的仮説とそれに対応するプロトコルは, この流れの最上流に位置しており, プロトコルに「エラー」があれば, 下流に位置する治療や解析の段階で「エラー」が増幅されつつ拡大再生産されることになってしまう。逆に, プロトコルの完成度が高ければ, あとはそれを遵守できれば, 正しい結論が導けるとも言える。プロトコルが, 「もっとも重要な品質管理ツール」であるとはそういう意味である。

多施設共同臨床試験のプロトコルを書く研究事務局(試験の責任者)にとっては, おのれがベストと信じるやり方で, すべての登録患者に同じ方針で治療を行い, その上で自分が興味をもっているclinical questionに対する答えを知りたいはずである。しかし, 現実には, 研究事務局がベストと考える治療法や治療変更規準, 支持療法を, 細部にわたるまで多施設のすべての担当医が同様にベストと考えるとは限らない。施設の担当医は, 医師法に基づく裁量権をもっており, 一般診療においては, 自らの責任において自らがベストと考える(あるいはその雇用責任者である施設の診療方針に基づいた)医療行為を行っている。多施設共同試験における治療とは, ある意味そうした医師の医療行為の裁量権に制限を課す特殊な医療行為であると言える。数十人, 時には数百人の他の医師の医療行為を,

一臨床医である研究事務局の責任において制限するのが多施設共同試験である。よって、治療法、治療変更規準、支持療法をプロトコルに記述する研究事務局には重大な責任があり、記述の不備により患者に生じた不利益は、第一義的には研究事務局の責任である。しかし、その責任は、一臨床医である研究事務局に負わせるには大きすぎる責任であることは論を待たない。そのために、多施設共同試験のプロトコルは、(JCOGでは、当該グループ内での十分な議論に基づいて作成され、さらにプロトコル審査委員会のpeer reviewによる審査・承認を得た上で)各施設の倫理審査委員会での第三者的審査に基づく医療機関の長の承認を得ることが不可欠なのである。プロトコルの審査・承認のプロセスは、研究事務局にとっては煩雑な事務的過程かもしれないが、実は研究事務局の医学的責任を分掌することで、研究事務局を保護する仕組みであることを忘れてはならない。

このことを踏まえれば、多施設共同試験のプロトコル作成にあたって、研究事務局が、事前に十分な議論を行って参加施設のすべての研究者の十分な合意を得ることや、プロトコルの記述が医学的に妥当であるだけでなく、論理的かつわかりやすくなければならないことは自明であろう。それは、責任を分掌してもらう“local user”, “central user”, “reviewer”すべての人に対する研究事務局の義務であり、また礼儀とも言えるであろう。JCOGでは研究支援部門でのプロトコル作成支援やプロトコル審査委員会審査で、誤字・脱字や日本語として不適切な表現の修正も行っているが、それはこうした考えに基づいている。日常診療だけでも多忙な施設の担当医に、プロトコルに規定したとおりの治療あるいは評価を行ってもらうのは容易ではないため、プロトコルが読みやすく、適切で、かつ十分に説得力のもつたものであることは最低限の必要条件と言えるだろう。

一方、試験開始後の品質管理活動としては、試験の進捗中に各施設においてプロトコル規定どおりの治療、評価が行われているかどうかを定期的にチェックし、不適格の可能性やプロトコル逸脱、重篤な有害事象などの問題を施

設側にフィードバックをかける「モニタリング」がなされる。つまり、プロトコルであるべき姿を示し、モニタリングによってそれらが遵守されていることを確認することが品質管理の大きな流れである。このモニタリングでは、個々の施設単位の問題だけでなく、多くの施設に共通する問題、すなわちプロトコルの記述の不備が見つかることも多い。いかに綿密に作成しても「完璧なプロトコル」はあり得ないため、プロトコルは試験開始後もモニタリングを通じて、常に改善・進歩していくものでもある。

参加する施設の研究者がプロトコルの内容をよく理解していないことや、プロトコル遵守の姿勢に欠ける研究者が中には存在することも事実であるが(明らかにプロトコルを読んでいると思われない場合もある)、これらを改善する努力は絶えず行わなければならない。そのためには、臨床試験を開始する前に全参加施設の研究者がプロトコルの内容を吟味した上で、それに合意することが必要であり、さらにそれに加えて、試験開始後は正しい結果を出すために、「自分たちの作った」プロトコルを遵守するという意識を高く保つことが、プロトコルを通じた品質管理のカギであると考えている。

3. 試験の意義を示す声明文

さて、プロトコルとは臨床試験の「取り扱い説明書」だと述べたが、もう一つ、プロトコルはその臨床試験を計画した「意義」を示す声明文的な役割も担っている。米国National Institutes of Health (NIH)のEmanuelら³⁴⁾は、臨床研究の8つの倫理要件を提唱しており(表1)、臨床試験のプロトコルでは、この中でも最初の4点、つまりSocial or scientific value, Scientific validity, Fair subject selection, Favorable risk-benefit ratioに十分に考慮されていることを示す必要があるとしている。一般的なプロトコルの章構成は次々回述べる予定であるが、これらのうちもっとも重要な「意義(Social or scientific value)」については、プロトコルの冒頭部分(JCOGでは2章「背景と試験計画の根拠」)に記載され、その臨床試験の「手順」については以降に続けられることが一般的である。これはreviewerにとっては、冒頭部分を読んで、この臨床試験

表1 臨床研究の8つの倫理要件(Emanuelらによる)

倫理要件	詳細要件
①Social or scientific value 社会的/科学的価値	診断・予防・治療の向上に貢献, 疾患・健康に有用な知識を得る, すでにある知識や無駄な重複ではない
②Scientific validity 科学的妥当性	一般的に認められた科学的方法論 適切な統計手法・正しいデータ
③Fair subject selection 適正な被験者選択	社会的弱者の保護・過大なリスクのある被験者の除外 利益を受ける集団とリスクを受ける集団が解離しない
④Favorable risk-benefit ratio 適切なリスク/ベネフィットバランス	リスクの最小化・ベネフィットの最大化 被験者のリスクに見合う被験者/社会のベネフィット
⑤Independent review 第三者審査	研究と利害関係をもたない独立した第三者による デザイン・対象・リスク/ベネフィットの評価
⑥Informed consent インフォームドコンセント	研究目的・方法・リスク・ベネフィット・代替治療の十分な説明 (information), 理解(comprehension), 自発同意(voluntariness)
⑦Respect for potential and enrolled subjects (候補者を含む)被験者の尊重	同意撤回の自由, プライバシー保護 開始後の新知見や研究結果の説明 継続的な被験者保護の監視
⑧Collaborative partnership 共同パートナーシップ	研究成果はコミュニティで共有 成果のみの搾取は防止 関係者は協調して研究を行う

の「意義」が理解できない場合や意義が乏しいと判断される場合には、その後の「手順」の記述を審査する価値がないということを意味する。また、local userにとっては、まず臨床試験の「意義」を十分に理解してから、実際の治療にあたるべきであることを意味している。このように、重要な「意義」の部分にSocial or scientific valueが十分かつ論理的に記述されている必要があるため、必然的に「背景」の記述は、ある程度の分量とならざるをえない(JCOGのプロトコルでは、通常10～15ページ程度)。

①～④の倫理要件を満たす内容が十分に記載されており、また説得力をもっているかどうか、第三者により審査されて承認が得られた(Emanuel 8要件のIndependent review)のち、個々の患者に「十分な説明」を行って、「理解」してもらった上で、「自発同意」によるインフォームド・コンセントを得て(Emanuel 8要件のInformed consent)、はじめて臨床試験による治療の開始が正当化される。

プロトコルは、研究者のみが理解できればよいという内部文書ではなく、第三者の審査の洗礼を受ける対外的文書、すなわち「他人が読む

ための文書」である。とかく臨床試験のプロトコルという、内部の取り決めのみが書かれた文書と捉えられがちであるが、他者に臨床試験の意義を理解させるためのツールでもあることは強く意識すべきである。

だれが書くか？

プロトコルの核となる部分の初稿は、もちろん実地臨床に携わり、その臨床試験におけるclinical questionをもっとも切実に感じている研究事務局が記載すべきである。具体的には、試験の意義、患者選択規準、治療、治療変更規準、安全性情報、評価スケジュールなどは臨床試験を立案した研究事務局自身が書くべき事項としてあげられる。しかし、実地臨床で多忙な臨床医が、時には100ページにもなるプロトコルを独力で作成することは不可能であり、また、多くの場合はじめて本格的なプロトコルを書く研究事務局が、プロトコル作成のノウハウをもたないことは当然である。JCOGでも古くは研究事務局自身がプロトコルを作成し、それがそのままプロトコル審査委員会へ提出されていたが、あまりにプロトコルの質のばらつき

が大きく、かつ標準化されていなかったため、データセンター長がプロトコル審査委員会提出前に、レビューもしくは直接部分執筆を行うようになり、その後プロトコルコーディネーター(医師)がデータセンター/運営事務局内での意見を取りまとめるとともに、自らも部分執筆を行う分担執筆体制へと変化してきた。そのため、現在ではプロトコル作成に関するノウハウは、JCOGの中でもプロトコルコーディネーターに蓄積されており、プロトコルコーディネーターは研究者と質疑応答を繰り返しながら不整合の修正、不足部分の追記を行って、研究者と二人三脚でプロトコルを完成へ近づけるとともに、これらのノウハウを最新のプロトコルに反映することで、プロトコルの質の向上ならびに標準化を図っている。

また、たとえば品質管理や品質保証、倫理や規制要件にかかわる定型的な部分や、データ管理や解析方法など、むしろcentral userであるデータセンター/運営事務局が記載すべき箇所も多い。現在、JCOGでは医師以外(non-MD)の研究支援部門スタッフが定型的な箇所を埋めて、書式を整えるとともに誤字・脱字や日本語表記のチェックを行っている。また、解析方法についてはデータセンターの統計家が記載し、当該試験担当のデータマネージャーがとくにデータ管理の部分に関してレビューを行っている。プロトコルコーディネーターはこれらのcentral userの意見をまとめ、研究事務局との窓口となってプロトコルを完成に導く役割を担っている。

現在、JCOGでは3名の医師がプロトコルコーディネーターを務めているが、米国の代表的な多施設共同試験グループであるSouthwest Oncology Group (SWOG: JCOGの兄貴分とも言える)のプロトコルコーディネーターは、すべてnon-MDである。全国的に医師不足が叫ばれている中、JCOGデータセンター/運営事務局の医師も限られているため、JCOGでもnon-MDのプロトコルコーディネーターの養成が必要と考えている。

高度な医学的知識が求められることから、non-MDがプロトコルコーディネーターとなるためには、一定期間の修練が必要であろうが、プロトコル作成のスピードアップや、安定したプロトコルの質の担保といったメリットも大きく、JCOGでもnon-MDのプロトコルコーディネーター養成に着手したところである。

おわりに

臨床試験が多くの組織で行われるようになり、それと呼応するように臨床指針の改正がなされるなど、プロトコルにかかわる規制要件も厳しくかつ細かくなってきた。必然的にプロトコルに書かれるべき事項は増え、倫理審査委員会の審査も厳しくなっており、「量」、「質」ともにこれまで以上の水準を求められている。多施設共同臨床試験のプロトコルは、もはや研究者ひとりの力で完成させられるものではなく、多施設・多職種の人々(これらの人々はユーザーでもある)が協力して完成させるものへと完全に移行したと言えるだろう。

今号では、プロトコルの「定義」、「機能」および「書き手」について概説を行った。次号では、引き続き「良いプロトコルの条件」や「プロトコルの標準化」などについて具体的な例をあげつつ解説を行う予定である。

文 献

- 1) Pocock SJ. Clinical Trials : A Practical Approach. Chichester : John Wiley ; 1983.
- 2) Piantadosi S. Clinical Trials : A Methodologic Perspective. 2nd ed. New York : John Wiley ; 2005.
- 3) Emanuel EJ, Wendler D, Grady C. What Makes Clinical Research Ethical? JAMA 2000 ; 283 : 2701-11.
- 4) Emanuel EJ, Wendler D, Killen J. What Makes Clinical Research in Developing Countries Ethical? The Benchmarks of Ethical Research. J Infect Dis 2004 ; 189 : 930-7.

腫瘍学における統計学

医学領域で用いられる統計学は一般に生物統計学とよばれ、腫瘍学（とくに臨床腫瘍学）の発展にも大きく寄与してきた。

臨床腫瘍学に関する臨床研究の例として、ここではKudoら（2008）により非小細胞肺癌患者を対象としてゲフィチニブ市販後に実施された観察研究を取り上げる。この研究では、EGFRチロシンキナーゼ阻害薬であるゲフィチニブ（gefitinib）と急性肺障害・間質性肺炎（ILD）の間の関連の評価を主目的の1つとし、ゲフィチニブ投与の有無によらず結果としてILDを発症した患者（これをケースという）1例に対してILDを発症しなかった患者（これをコントロールという）4例を選択してそれぞれについて詳細なデータが収集された。

オッズ比

この研究の結果を【表1】に示した。もしもゲフィチニブ投与とILD発症の間に関連がないならば、ケースにおけるゲフィチニブの投与割合はコントロールにおけるそれと同程度になることが期待できる。ケース・コントロール研究では、この関連を評価・定量するために一般にオッズ比とよばれる指標を用いる。以降、オッズ比とは【表2】で定義されるものをさす。このオッズ比はゲフィチニブ投与とILD発症に関連がない場合に値が1となり、ゲフィチニブ投与によってILD発症が増えるのであれば1よりも大きな値となる。前述のデータから【表3】のようにオッズ比を求めると2.35となる。

ここで求めたオッズ比の値をもって、ゲフィチニブ投与によりILD発症が増えた結論づけでもよいであろうか。答えは「No」である。その理由として、①観察された結果はまったくの偶然であり、真実としてはゲフィチニブ投与とILD発症の間には関連がないかもしれないこと（偶然性）、②病状が進んだ患者はそもそもILDを発症しやすく、そのような患者にゲフィチニブが投与される傾向があるかもしれないこと（交絡）、③ILDの診断は難しいことからケースも真実にはILDではないかもしれないこと（誤分類）があげられる。少なくともこの3つの可能性を十分に排除できない限り、前述の結論は必ずしも適切でないといえる。以下ではこの例を題材にして①～③の順に統計学の枠組で議論を行ってみることにする。

95%信頼区間

①は観察結果に含まれる偶然性を評価することの重要性を表している。真実としては関連がないにもかかわらず偶然的な差（誤差）によって2.35と観察された可能性が高いのであれば、前述の結論を出すのは早計であろう。

統計学を用いれば、観察結果に含まれる偶然性を定量化し、これを範囲・区間として明示することも可能である。医学研究では、この区間として慣習的に95%信頼区間を用いることが多い。95%信頼区間とは、仮に何度も研究を繰り返して、それら研究ごとに95%信頼区間を求めたとした場合に正しい値（真値）を100回中95回はその区間内に含むように構成したものである。ただし実用上は、「求めた95%信頼区間内に真値を含む確率が95%である」との解釈を行ったとしても大きな問題とはならない。

適切な統計手法を用い、【表1】のデータに基づいてオッズ比の95%信頼区間を求めると1.56～3.52となる。偶然性を考慮して構成された区間にオッズ比1を含まないため、まったくの偶然のみで2.35が観察された可能性は十分に低いと判断できるであろう。

交絡と層別解析

次に②は交絡（confounding）という現象を表しており、この交絡の原因となる因子を交絡因子という。病状が進むとゲフィチニブが投与される傾向があるとともに、ゲフィチニブ投与の有無によらずILDを発症しやすくなるという構造【図1】がある場合、病状が交絡因子となっており、ゲフィチニブ投与とILD発症の間に観察された関連はこの病状を介した見かけ上のものであり、必ずしも真にこのような関連があるわけではない。

適切な統計的手法を用いれば、この交絡の影響を減らすことができる。例えば、【図1】の構造が正しいのであれば病状が進んでいない患者のみに限って先の【表1】のように結果をまとめたとした場合、病状が同程度である患者の間ではこの影響が小さくなることから交絡の影響を減らすことができるであろう。一方で病状が進んだ患者のみに限って評価を行っても同様のことがいえる。このようなアプローチを層別解析といい、交絡を考慮・調整する目的としては用いられる。層別解析のほかにも、これを一般化したものに相当するロジスティック回帰など統計モデルを用いて調整を行うことも一般的である。Kudoら（2008）は、モデルを用いて年齢、全身状態、罹病期間、心疾患、合併症、喫煙状況などの交絡因子となりうる主なものを調整して求めたオッズ比として3.23（95%信頼区間1.94～5.40）も報告している。この結果に基づけば、これらを調整したとしても、ゲフィチニブ投与とILD発症の間の関連は否定できないと判断できるであろう。

表1 ゲフィチニブ投与の有無とILD発症の関連

ケース コントロール	ゲフィチニブ投与		計
	有	無	
79	43	122	
252	322	574	

表2 オッズ比の算出方法

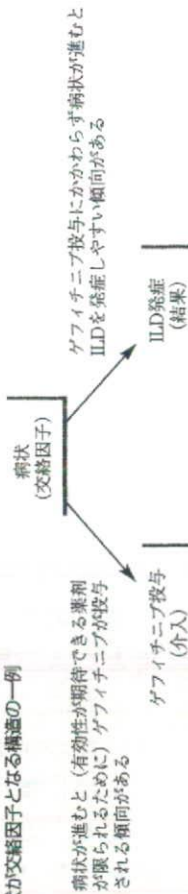
ケース コントロール	ゲフィチニブ投与	
	有	無
a	b	
c	d	

オッズ比 = $\frac{a}{b} / \frac{c}{d} = \frac{ad}{bc}$

表3 ゲフィチニブ投与の有無とILD発症の関連に関するオッズ比

ケース コントロール	ゲフィチニブ投与		オッズ比 = (79/43) / (252/322) = 2.35 オッズ比の95%信頼区間: 1.56~3.52
	有	無	
79	43	122	
252	322	574	

図1 病状が交絡因子となる構造の一例



感度・特異度

最後に③は誤分類とよばれる現象を表す。ここで誤分類とは真にはILDを発症していない患者がILDと診断されること（偽陽性）、または真にはILDを発症している患者がTLDと診断されないこと（偽陰性の両方）をさす。誤分類が存在する場合、正しいオッズ比とは異なる偏り（バイアス）のある結果が得られる可能性が高まることから注意が必要である。

誤分類の程度を表す指標として感度 (sensitivity) と特異度 (specificity) がある。ここで感度とは真にILDを発症している患者を正しく診断できた割合、特異度とは真にILDを発症していない患者を正しく診断できた割合にそれぞれ対応する【表4】。感度、特異度ともに値が大きいかほどその診断能力が高いこと、すなわち誤分類が少ないことを表す。

適切な統計手法・研究デザインを用いれば誤分類による影響を減らすことも可能である。高度な統計的考え方を必要とするために詳細は省略するが、Kudo et al. (2008) では誤分類を考慮した解析の結果も提示してその評価を行っている。

ここでは1つの研究例を用いて解説を試みたが、統計学が臨床腫瘍学の発展に大きく寄与してきたことが少しでも伝われば幸いです。エビデンスに基づく医療 (EBM) を実践する観点からも、適切な統計手法を用いてデザインされた研究、およびそれらのデータを解析した結果に対して適切な解釈が行える素養を身につけることは重要であり、その重要性はますます高まっています。また実際に一研究者として研究を行う場合にも統計的考え方を念頭におき、適切に研究を行うことができます。求められています。

表4 真のILD発症と診断されたILD発症

		実際に診断されたILD発症	
		有	無
真のILD発症	有	x	y
	無	z	w
		感度 = $x / (x + z)$	
		特異度 = $w / (z + w)$	

Basic Point public health

●罹患率
疾病に新たにかかることを罹患といひ、特定の集団においてある期間に発生した特定の疾患の新規患者数を、その疾病に新たにかかるリスクにあった人口で除したものを罹患率 (incidence) という。特定の集団においてある時点で特定の疾患を有する者の割合を表す有病率 (prevalence) とは異なる。予防対策に用いる指標としては有病率よりも罹患率を用いるほうが適切である。

Basic Point public health

●死亡率
特定の集団において単位時間における死亡者数の割合を死亡率 (相対死亡率) という。単位時間としては1年が用いられることが多い。集団間で年齢構成が異なるような場合、年齢構成による交絡のため、粗死亡率では適切な比較が行えない。例えば、昭和60年と平成15年の間で死亡率の比較を行う場合がこれに該当する。この場合には、直接法により率を標準化 (standardization) することで得られる年齢調整死亡率や、間接法により率を標準化することで得られる標準化死亡比 (standardized mortality ratio: SMR) を用いることが一般的である。

Level Up View

●がん対策基本法 (法号高号：平成18年法律第98号)
がん対策基本法は第164回通常国会において提出された議員立法であり、2006年6月6日成立し、2007年4月1日に施行された法律である。これを期に、わが国は更なるがん対策を推進することとなった。がん対策基本法の第一条には厳しい現状認識とともにその目的が述べられている【表A】。
がん対策基本法には、①がんの予防と早期発見の推進、②がん医療の均てん化の促進、③がん研究の推進が本法の本柱として明示されており、行政としてこれらを推進していくことが期待される。

表A がん対策基本法 第一条

この法律は、我が国のがん対策がこれまでの取組により進展し、成果を収めてきたものの、なお、がんが国民の健康による死亡の最大の原因となっている等が、国民の生命及び健康にとって重大な問題となっている現状にかんがみ、がん対策の一層の充実を図るため、がん対策に関し、基本理念を定め、国、地方公共団体、医療保険者、国民及び医療等の責務を明らかにし、並びにがん対策の推進に関する計画の策定について定めるとともに、がん対策の基本となる事項を定めることにより、がん対策を総合的かつ計画的に推進することを目的とする。

○文献

- 1) Kudo S, et al.: Interstitial lung disease in Japanese patients with lung cancer: a cohort and nested case-control study, Am J Respir Crit Care Med, 177:1348-1357, 2008.

Self Check

- 95%信頼区間は偶然性を考慮して結果を提示する目的で用いられる。
- 交絡が存在する場合、観察結果には偏り (バイアス) が含まれる。
- 感度、特異度は診断能力の指標として用いられる。
- 感度とは、真には疾患をもつ患者に対して正しく診断できる確率である。
- 特異度とは、真には疾患をもたない患者に対して正しく診断できる確率である。

(吉村健一)



臨床試験の方法が進歩し、適切なエビデンスを提供できるようになり、計画された試験の結果も世界的に数多く公表されている。公表結果を正しく理解するには、情報の受け手にも生物統計学に関する適切な理解が不可欠である。

概要

標準治療に比べて、新治療が優れるかを検証する目的で行われる試験を検証的試験という。検証的試験はランダム化比較試験(RCT)として実施されることが一般的であり、 α エラーを犯す確率が十分に小さくなるように有意水準が定められる。その計画段階では統計的有意性と臨床的有意性のあいだで乖離が生じないように、臨床的最小有意差(minimum clinically significant difference)に對する検出力を適切とする対象者数を定めることが肝要である。近年では、倫理的観点などから試験途中で中間解析が行われるようになっており、早期中止によって早期に結果公表されることがある。

ランダム化比較試験

対象者をコントロール(対照)群を含む2群以上の試験治療群にランダムに割りつけることで、医学的介入(治療法)を評価する試験デザインである。RCTを実施するにあたっては、少なくとも裏に列挙した事項を一般に考慮する必要がある。介入をランダムに割りつけることをランダム化と呼ぶ。ランダム化は現代の推計統計学の確立者の1人であるSir Ronald Aylmer Fisherが1930年代に著書「実験計画法」で定式化した統計的方法である。RCTにおいてp値を算出する際の確率的基礎であり、p値を算出する際にはランダム化した事実以外に仮定を要さない。

統計的有意性と臨床的有意性

統計的有意性は観察された群間差などの結果が偶然性を超えるものであることを意味し、必ずしもそれがそのまま臨床的にも有意な差が存在することに対応しない。統計的有意性は当該試験の対象者数に大きく依存するためである。例として、血圧を0.5 mmHg低下させる降圧薬はその有害事象も考慮に入れると臨床的に有意でないことは明白であるが、対象者数を多くすればするほどこの些細な差に対しても統計的有意性を示すことが容易となる。反対に対象者数を少なくすればするほど、真に強い作用をもつ降圧薬であろうとも、その統計的有意性を示すことは困難となる。対象者数が少なすぎるために臨床的に有意な差を検出できない試験(low power trial)は倫理的にも問題があり実施すべきでない。統計的有意性と臨床的有意性のあいだで乖離が生じないようにするために、試験計画時に適切な対象者数を定めることが肝要となる。そこでは、試験目的に応じた臨床的最低有意差を設定した上で十分かつ過大でない検出力となるよう対象者数を決定する。

用語解説——有意水準

多くの領域で慣習的に(片側検定の場合には)片側有意水準5%あるいは(両側検定の場合には)両側有意水準5%が用いられている。両側有意水準5%は、片側検定で検み替えを行うと片側有意水準2.5%に対応する。

用語解説——p値

観察された結果の「帰無仮説が真という仮定の下での不自然さ」を数値化したもの。0に近づくほど不自然さを表す。仮説検定ではこの値が有意水準を下回るか否かによって帰無仮説を否定するか否かを判断する。

表 ランダム化比較試験を実施する際に考慮すべき事項

①	明確な臨床仮説に対する仮説検証的試験として実施
②	偏りなく測定できる明確なプライマリーエンドポイントを設定
③	プライマリーエンドポイントの性能性や解釈を補う目的のサセンドポイント/セカンダリーエンドポイントを設定
④	研究目的に応じた適格標準により研究対象集団を定義
⑤	倫理的観点、科学的観点から適切なコントロール群(例:プラセボ、標準治療)を設定
⑥	倫理的観点(インフォームド・コンセント、試験中の安全性モニタリングなど)

α エラー

本当は帰無仮説が正しいにもかかわらず、帰無仮説を否定するという判断を犯す誤りである。一般に、 α エラーを犯す確率は、試験デザインに沿った適切な統計解析法を用いることで事前に定めた有意水準以下となることが保証されるが、適切な試験デザインおよび統計解析法を用いない場合には必ずしも制御されない。 α エラーを犯す確率は事前に設定した有意水準に一致し、観察された結果に基づいて算出されたp値には依存しない。その定義から、試験後に帰無仮説を否定できなかつた場合(つまり統計的有意性が示されなかつた場合)にはこの種のエラーは生じない。

中間解析

試験の正式な完了以前に、試験の早期中止の判断を意図して行う解析である。国際的ガイドライン「臨床試験のための統計的原則」においては、「試験の正式な完了以前に、有効性または安全性に関して試験治療群間を比較することを意図して行われるあらゆる解析」として定義されている。これに従えば、臨床試験の途中で群間比較を行うことを中間解析といい、有効性に関する群間比較のみに限らず、安全性に加えて中間解析を実施する場合、中間解析を計画した試験デザインを採用する場合、最終解析に加えて中間解析を実施するという意味で計2回以上の群間比較を同一試験内で行うことになるため、多重性と呼ばれる問題が生じる。適切な補正を伴わずに中間解析結果を提示することは、 α エラーを犯す確率を一定水準内に制御できないため不適切である。これを補正するために、中間解析法と呼ばれる統計手法を用いることができる。また、中間解析で早期中止された試験で観察される群間差は多重性によって過大評価傾向があるため、その解釈には注意を要する。中間解析法により補正されるものは一般に α エラーを犯す確率のみであり、過大評価された群間差までは補正されない。

Recommended Readings

- Armitage P, Berry G: *Statistical Methods in Medical Research*, 3rd ed. Blackwell, 1994(第3版) 第5章「区別化するための統計的方法」サイエンス出版社, 2001
- 黒川均典: *新臨床試験—がん薬物療法専門医のための*。南江堂, 2006, pp.133-137
- 佐藤健一: *新臨床試験—がん薬物療法専門医のための*。南江堂, 2006, pp.138-142
- 吉村健一: *新臨床試験—がん薬物療法専門医のための*。南江堂, 2006, pp.127-133
- 吉村健一: *腫瘍内科1*: 144-146, 2007

関連事項

- 徳の臨床試験のデザインとその特徴 ▶▶▶ 46頁
- レシスト ▶▶▶ 48頁
- NCI-CTC ▶▶▶ 50頁
- バイオマーカー研究 ▶▶▶ 108頁

第26回臨床薬理阿蘇九重カンファレンス IRBの現状と課題

原著

福岡県における研究倫理審査委員会の実態調査

—IRBメンバー教育システムの構築を目指して—

笹栗 俊之 柴田 智美 上口 愛 白石富美恵
 三輪 宜一 高橋 富美 森本 幸生

九州大学大学院 医学研究院 生体情報科学講座 臨床薬理学分野

Questionnaire survey on the institutional review boards in Fukuoka aiming at the construction of the board member education system

Toshiyuki Sasaguri Tomomi Shibata Megumi Ueguchi Fumie Shiraishi
 Yoshikazu Miwa Fumi Takahashi-Yanaga Sachio Morimoto
 Department of Clinical Pharmacology, Faculty of Medical Sciences, Kyushu University

Abstract

We investigated the present conditions of the institutional review board (IRB) to collect basic information for building a board member education and training system, assuming Fukuoka prefecture a model district. A questionnaire survey was performed for all IRBs in the prefecture.

For 173 committees possibly having a duty of the IRB which we understood in a preliminary investigation of June, 2006, we performed this investigation in September of that year and obtained an answer from 163 committees (recovery rate 94.2%). However, based on their answers, we selected 137 committees of these to analyze their answers. The number of IRBs which reviewed only trials for approval of drugs/equipment (group A) was 71, that of IRBs which reviewed only studies except the trials for approval (group B) was 36, and that of IRBs which reviewed both (group C) was 30.

The IRBs which excluded the director or president of the institution were 77.5% in group A, but it was only 36.1% in group B. IRBs with the possibility that the director or president became a chairperson were 22.5% (5.63%, actually chairperson) in group A and 58.3% (36.1%, actually chairperson) in group B. There was a remarkable difference in the number of protocols reviewed. Some IRBs reviewed almost 100 new protocols per year, whereas 19 IRBs did not review any protocols in 2005. For 33% of IRBs in group B and 20.0% of those in group C some protocols were not able to be reviewed, even if they were studies that should be reviewed. The IRBs meeting almost regularly were 74.6% in group A, but it was only 33.3% in group B. The deliberation time was under one hour for 70.0% of the IRBs in group A, but, on the other hand, 10.3% of the IRBs in group C spent three to four hours. On the administration side, there were many IRBs in which the schedule adjustment was often difficult, in group B in particular. The IRBs which allowed the investigator concerned to be present at the time of voting were 14.1% in group A, whereas those were 30.6% and 41.4% in group B and group C, respectively. Only four protocols were rejected in trials for

approval of drugs/equipment, and it was only nine in other studies in 2005. The reason for rejection was often that “the risk exceeded benefit” in the trial for approval, whereas in other studies it tended to be a problem in the purpose or methods. There were few IRBs which opened review results to the public. There were extremely few IRBs which used original review criteria and some kind of procedural manuals. “Adequacy of the placebo use” and “decision about the study with little benefit for subjects” were often raised as the points that board members felt difficult to judge. To “make practical criteria for IRB” was the most often raised proposal for solving this difficulty. For 90% of IRBs, there was no opportunity for education or training of the members, and around 80% of IRBs answered that they wanted to have such opportunities.

It seemed that we were able to detect more than 90% of the IRBs in the prefecture by this investigation, and valuable data were obtained to develop an education and training program for the board members. Most of the IRBs needed education and training, and many other problems were highlighted. Based on these results, we wish to network IRBs in the prefecture and perform an effective instructional activity.

Key words

research ethics committee, institutional review board, questionnaire survey, Fukuoka prefecture, education and training system

Rinsho Hyoka (Clinical Evaluation) 2008 ; 36 : 393-419.

はじめに

被験者の人権を守り安全を確保するため、人を対象とする医学研究（臨床研究）は倫理審査を受けなければならない。今日では、「医薬品の臨床試験の実施の基準」（GCP）や「臨床研究に関する倫理指針」等により、臨床研究は、研究倫理委員会（Research Ethics Committee：REC）により審査されることが義務づけられている。日本の場合、RECは、研究実施施設に設置されるのが基本となっており、施設内審査委員会（Institutional Review Board：IRB）と呼ばれることが多い¹⁾。

IRBが臨床研究を承認するということは、一般社会にその適切性を保証することを意味し、その役割は重大だが、倫理審査のあり方には様々な問題が今も残されている。

その中でも最大の問題は、審査の質が保証されていないことであろう。日本の法規やガイドラインは、IRBを設置して審査することは、たしかに求めている。しかし、何に基づいてどう判断したらいいのか、つまり審査の基準や方法についてはほとんど触れず、各々のIRBに任せきりである。行き当たりばったりの審査では、質が保たれるはずはない。審査の質が確保できないもう一つの理由は、IRB委員に特別の資格が必要ないばかりか、委員に就任した後も教育や研修を受ける機会がほとんどないことである。委員の質が確保できなければ、審査の質も望みようがない。

すなわち、審査の質を保証するには、第一に、審査の基準となる倫理原則を定め、それに基づく標準的な審査方法を作成すること、第二に、IRB委員（および事務担当者）のため、倫理審査に関する教育・研修の機会を設けることが必要である。

しかしながら、これらを実現するには、IRBの実態を十分把握した上で、IRBどうしを連絡する横のつながり（ネットワーク）を作ることが重要と思われる。なぜなら、倫理原則を定めたり審査方法を標準化したりするには、倫理審査に関わる大勢の人たちが集まり、知恵や意見を出し合うこ

とが最も効果的だと思われるからである。また、ネットワークがあれば、教育・研修のシステムを作るのも容易になるからである。

ネットワークを作るには、どこにどのようなIRBがいくつ存在するか把握する必要がある。ところが、今のIRBには登録制度がないため、IRBの存在を系統的に把握することは非常に難しい。特に全国規模で全IRBを把握することは事実上不可能に近い。

そこで今回、福岡県をモデル地区とし、県内の可能な限り全てのIRBを探索し、これらを対象として詳細な実態調査を行い、IRBの問題点を抽出するためのデータベースを作成することにした。これにより、どのようなIRBがどこにあるか把握できるとともに、活動実態とその問題点がわかり、また、ネットワークを作るための情報基盤が得られると予想された。

方 法

1. 用語の定義

本論文では、IRBという用語を「臨床研究の倫理審査を行う委員会」という意味で用いている。その意味では、本来ならば、より普遍的な用語であるResearch Ethics Committee (REC)（日本語では「研究倫理委員会」や「研究倫理審査委員会」など）を用いる方が望ましい。なぜなら、IRBとはInstitutional Review Board (施設内審査委員会)の略で、「施設単位に設置するREC」を意味しており、設置様式に由来する名称に過ぎず、普遍性がないからである。しかしながら、日本では、RECという用語があまり一般化していないこと、現実的にはRECは施設単位に設置されることが多いため、IRBと呼んでも国内ならほとんど問題ないこと、また、RECという用語を用いると「治験以外の一般臨床研究の審査を行う委員会」であるかのような誤解が生じる可能性があること（これ自体問題であり、後ほど取り上げる）などから、本論文の用語では、日本におけるRECをIRBという用語で統一した。

もう一つ問題となったのは、いわゆる「親委員会」と「専門部会」の別である。今日、特定の研究課題（例えばゲノム解析研究）の審査を専門的に行う「専門部会」が、「親委員会」の下にしばしば設置されている。しかしながら、これらを「親委員会」と同列に解析するのは適切ではないと考えられたため、今回の解析からは除くことにした。ただし、調査は同様に行い、後に構築する予定のIRBのネットワークには加えることにした。

2. 調査対象

福岡県全域をモデル地区とし、県内の全IRBを調査対象とした。ただし、現在、IRBには登録制が存在しないので、全てのIRBを正確に把握することは非常に難しいと予想された。もちろん、全ての医療機関と研究機関を対象として調査すれば、ほとんど全てのIRBを把握できると思われたが、診療所（病床数19床以下の医療機関）を含めた調査は費用の不足により困難であった。

そこで今回は、可能な限り全てのIRBを調査するための現実的方法として、県内の全ての病院および医学研究機関（約500施設）に調査票を送っ

た。臨床試験を専門とする医療機関以外でIRBが診療所に設置されることはまれなので、おそらく大きな問題とはならないと考え、臨床試験を専門に行っている機関など一部の例外を除き、診療所には調査票を送らなかった。

3. 予備調査

福岡県の全IRBのリストを作成するため、2006年6月から9月にかけて、上記の約500施設を対象とし、予備調査票（Table 1）を用いて、どこにいくつのIRBが存在するか調査した。回答が得られない場合、最大3回まで繰り返し依頼を行ったが、9月中旬で締め切り、それまでに回答が得られなかった施設にはIRBはないものと判断した。

4. 本調査

予備調査により、IRBの役割を有する可能性があると思われた委員会を対象として、2006年9月から2007年2月にかけて、全68問から成るアンケート調査「IRBメンバー教育研修プログラムの開発を目的とするIRBの実態調査」（資料・文末）を行った。

Table 1 予備調査票

<p>問1. 貴施設に、医学・医療上の倫理的な問題を検討する委員会がありますか？あるいは設置の予定はありますか？（たとえば、「倫理委員会」や「治験審査委員会」のような委員会のことです）</p> <p style="text-align: center;"> <input type="checkbox"/> ある <input type="checkbox"/> ない </p> <p>問2. 「ある」とお答えになった方へ、委員会の名称をお書き下さい。（複数ある場合は全てお書き下さい。専門部会などがある場合もお書き下さい。不足なら余白に）</p> <p>① _____</p> <p>② _____</p> <p>③ _____</p> <p>これらの委員会のうち、IRBに相当する委員会（治験や、その他の臨床研究の倫理審査を行う委員会）はありますか？ある場合は、その番号（①、②など）を全てご記入下さい。</p> <p style="text-align: center;"> <input type="checkbox"/> ある（番号 _____） <input type="checkbox"/> ない </p> <p>問3. IRBメンバーのための教育研修システムができれば、参加を希望されますか？（施設にIRBがなくても、関心があれば参加できます）</p> <p style="text-align: center;"> <input type="checkbox"/> はい <input type="checkbox"/> いいえ <input type="checkbox"/> わからない </p>
--

客観的な質問にはIRB事務局の担当者等が答えてもよいことにしたが、やや主観的な質問への回答を含め、最終的には当該IRBの委員長が確認することとした。まず郵送で回答を依頼したが、書面回答だけでは不明確な点が多かったため、ほとんど全ての施設に調査員が直接出向き、委員長や事務局担当者等から聴き取りを行い、データを完全なものに近づけた。

5. 倫理面への配慮

調査項目に回答者の氏名、IRB委員長の氏名を記入する欄があるため、データは研究室から持ち出さない、ファイルは限定した専用コンピューターにしかコピーしないなど、個人情報保護を徹底した。

結果

2006年6～9月に行った予備調査の結果、IRBとしての役割を有する可能性が高い委員会を、福岡県内に173確認できた。これらを対象として9月から本調査を行い、このうち163委員会から回答を得た(回収率94.2%)。残りの10委員会のうち、5委員会は回答を拒否し、あとの5委員会は、最終締め切り日の2007年2月末日までに回答がなかった。

ただ、アンケートの回答を見たり、直接尋ねたりした結果、すべてがIRBに該当する委員会ではないこと、また、同列には扱えない委員会も含まれていることがわかり、次の委員会を今回の解析対象から除外した。①いわゆる「親委員会」の下位に位置する「専門部会」、②病院倫理委員会(臨床倫理委員会)*¹に該当する委員会、③活動実態がない委員会(過去にはあったが最近数年はない委員会、および、現在は名前だけで実体はこれから作る予定という委員会)。その結果、137委員会が解析対象となった。

当初は、すべてのIRBを一括して解析しようとしたが、どのようなIRBに問題があるのかをより明確にするため、問20への回答に基づき、IRBを審査対象の異なる3つのカテゴリー(A群、B群、C群)に分けて解析することにした(Table 2)。これにより、解析対象とした137のIRBは、A群71委員会、B群36委員会、C群30委員会に分類された。

以下、設問(資料・文末)の順に調査結果の要点を述べる。

1. 委員会の概要について(問1～4)

問1と問2は委員会の名称と委員長の氏名を尋ねる問いであり、結果は省略する。設置された時期については、いずれのカテゴリーでも「2001～

Table 2 審査対象によるIRBの分類

A群：治験のみを審査するIRB

正確に言えば、治験の他、製造販売後調査および製造販売後臨床試験を審査対象とするIRB、すなわち、薬事法に基づくGCPに準拠して実施する医薬品・医療機器の承認申請を目的とする研究を審査する委員会であるが、簡便のため「治験のみを審査する」と表現した。多くは企業主導の開発研究だが、医師主導の治験も含まれる。ただし、この群に分類したIRBでも、審査対象のすべてがGCPに準拠するとは限らず、治験には該当しないが企業から委託されて行う研究を例外的に審査するとしても、この群に含めた。

B群：治験以外の研究のみを審査するIRB

GCPに準拠しない、A群が対象とする研究以外の研究(一般臨床研究)を審査するIRB。

C群：どちらも審査するIRB

対象とする研究がA群とB群の両方にまたがり、GCPに準拠する研究と準拠しない研究のどちらも審査対象とするIRB。

*¹ 臨床研究ではなく、診療上の倫理的問題を扱う委員会。

2005年」という回答が最も多かった。ただし、1995年以前に設置された委員会は、A群では21.1% (15/71^{*2}) だったが、B群では27.8% (10/36)、C群では53.3% (16/30) に上った (問3, Fig. 1)。委員会の任務について尋ねた問4は、その委員会がIRBの定義を満たすかどうかを確認する目的で設けた。例えば、「臨床研究の審査」にはチェックがなく、他の項目のみにチェックがある委員会は、先に述べた「除外基準」の②、すなわち病院倫理委員会や臨床倫理委員会などと呼ばれる委員会の可能性があると考え、直接問い合わせて確認した上で解析対象とするかどうか判断した。

2. 委員会のメンバーについて (問5～19)

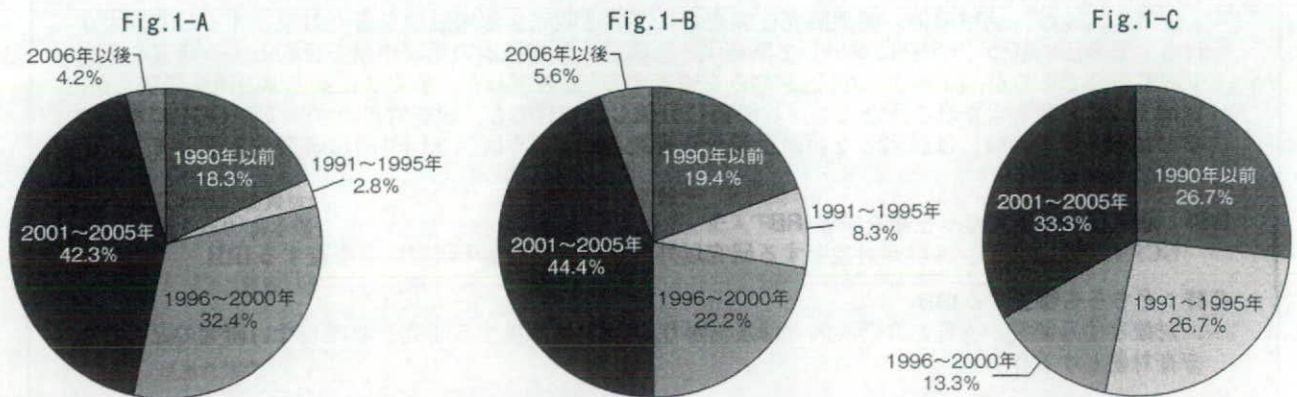
構成人数は、全体としては10.6±3.03名 (うち女性2.05±1.39名、施設外委員2.08±1.57名)、A群では9.79±2.62名 (うち女性2.32±1.44名、施設外委員1.96±1.07名)、B群では11.1±3.93名 (うち女性1.89±1.09名、施設外委員1.97±1.30名)、C群では10.9±2.43名 (うち女性1.93±1.55名、施設外委員2.30±2.60名) であった (問5, Table 3)。構成員は、いずれの群でも「医師」が最も多く、全体で4.72±2.52名、A群3.92±2.09名、B群5.47±3.24名、C群4.77±2.14名を占め

ていたが、「臨床薬理学会認定医」は少なく、全体で0.09±0.43名、A群0名、B群0.03±0.32名、C群0.23±0.68名に過ぎなかった。「薬剤師」は、全体で1.10±0.77名、A群1.24±0.80名、B群0.75±0.49名、C群1.30±0.89名、「臨床薬理学会認定薬剤師」は、全体0.07±0.35名、A群0.03±0.24名、B群0.08±0.48名、C群0.10±0.40名であった。「一般の立場を代表する者」の人数は、全体1.65±1.10名、A群2.01±1.03名、B群1.42±1.19名、C群1.53±1.14名であった (問6, Table 3)。「一般の立場を代表する者」の職業としては「会社員」がいずれの群でも最も多く、全体で27.0% (37/137)、A群31.0% (22/71)、B群13.9% (5/36)、C群33.3% (10/30) を占めたが、「その他」とされた者の職業も含めると、当該施設の職員が多くを占めると考えられた (問7)。これに一致して、その人を選んだ理由としては「病院の職員である」 (問8) が最も多かった。適任者を探すのは「どちらかと言うと難しい」と答えたIRBがいずれの群でも最も多かった (問9)。

構成員から施設長 (病院長、学部長など) を除外しているIRBは、全体で64.2% (88/137) に過ぎず、含む可能性がある委員会が35.8% (49/137) あり、実際含んでいる委員会が22.6% (31/137) もあった (問10)。除外しているIRBをカテゴリー

Fig. 1 設置時期

資料の問3「委員会が設置されたのはいつですか」への回答による。A, B, Cは、Table 2のA群, B群, C群を示す。



*2 分数の分子は該当する委員会の実数、分母は有効回答総数。以下同様。

別に見ると、A群では77.5% (55/71)、C群では66.7% (20/30) だったが、B群では36.1% (13/36) に過ぎなかった (Fig. 2)。しかも、施設長が委員長になる可能性のあるIRBが、全体で32.8% (45/137) あり、カテゴリー別に見ると、A群では22.5% (16/71)、C群では26.7% (8/30) だったが、B群では58.3% (21/36) もあった (問11, Fig. 3)。現に委員長であるIRBは、全体で15.3%

(21/137)、A群では5.63% (4/71)、C群では13.3% (4/30) で、B群では36.1% (13/36) にも上った (Fig. 3)。それにも拘わらず、第三者の立場に立った審査が「十分できている」あるいは「おおかたできている」と答えたIRBは、全体で89.1% (122/137) を占め、A群が87.3% (62/71)、C群が90.0% (27/30)、むしろB群が最も多く91.7% (33/36) を占めた (問12)。

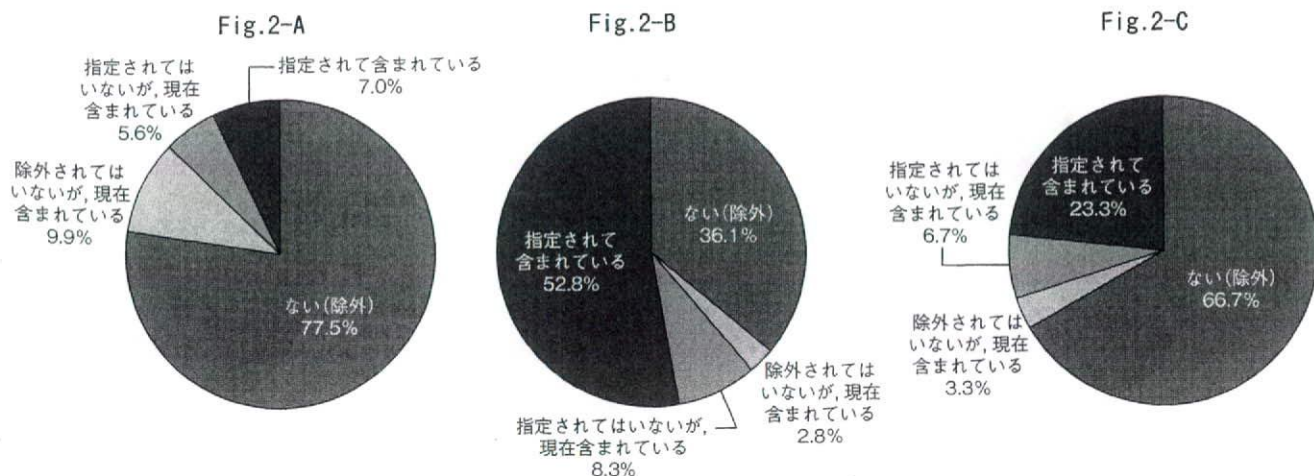
Table 3 構成メンバー

資料の問5および問6への回答による。単位は「人」。

	全体	A群	B群	C群
全構成員数	10.6 ± 3.03	9.79 ± 2.62	11.1 ± 3.93	10.9 ± 2.43
医師	4.72 ± 2.52	3.93 ± 2.09	5.47 ± 3.24	4.77 ± 2.14
臨床薬理学会認定医師	0.09 ± 0.43	0	0.03 ± 0.32	0.23 ± 0.68
歯科医師	0.29 ± 1.54	0.13 ± 0.95	0.28 ± 1.99	0.47 ± 2.10
薬剤師	1.10 ± 0.77	1.24 ± 0.80	0.75 ± 0.49	1.30 ± 0.89
臨床薬理学会認定薬剤師	0.07 ± 0.35	0.03 ± 0.24	0.08 ± 0.48	0.10 ± 0.40
看護師	1.16 ± 0.89	1.23 ± 0.86	1.25 ± 0.78	1.00 ± 1.12
その他の医療職者	0.54 ± 0.84	0.48 ± 0.70	0.53 ± 1.16	0.60 ± 0.79
統計学の専門家	0.02 ± 0.23	0.01 ± 0.18	0.03 ± 0.32	0.03 ± 0.25
法の専門家	0.43 ± 0.64	0.24 ± 0.56	0.64 ± 0.60	0.40 ± 0.68
倫理学者	0.06 ± 0.37	0.01 ± 0.19	0.03 ± 0.32	0.13 ± 0.58
心理学者	0.01 ± 0.19	0.01 ± 0.19	0.03 ± 0.33	0
その他の専門家	0.25 ± 0.71	0.17 ± 0.60	0.25 ± 0.63	0.33 ± 0.92
宗教家	0.03 ± 0.35	0.04 ± 0.40	0.03 ± 0.32	0.03 ± 0.27
一般の立場を代表する者	1.65 ± 1.10	2.01 ± 1.03	1.42 ± 1.19	1.53 ± 1.14
その他	0.50 ± 1.00	0.41 ± 0.89	0.58 ± 1.20	0.50 ± 0.99

Fig. 2 施設長が委員に含まれる可能性

資料の問10「施設長や委員会が設置されている部局の長（病院長、学部長など）が、委員会メンバーに含まれることがありますか」への回答による。A, B, Cは、Table 2のA群, B群, C群を示す。



委員の報酬については、「施設外委員にのみある」と答えたIRBがいずれの群でも最も多く、A群で81.4% (57/70), B群で55.6% (20/36), C群で53.3% (16/30) を占めた (問13). 委員の任期は、「決まっている」と答えたIRBがいずれの群でも多く、平均年数は、A群1.83±0.47年, B群1.88±0.34年, C群1.71±0.46年だった (問14). 委員長の選出方法は、いずれの群でも「施設長の指名」が最も多く、A群で50.7% (36/71), B群で30.6% (11/36), C群で53.3% (16/30) を占めた (問15). 成立に必要な出席者数についてはA群とC群では「過半数」、B群では「3分の2以上」が最も多かった (問16). 関連する倫理規範 (綱領) の理解度については、「ほとんどすべての委員が理解している」または「理解している委員の方が多い」と答えたIRBが、A群で71.4% (50/71), B群で82.9% (29/35), C群で83.3% (25/30) だった (問17). 委員は審議資料を事前によく読んでいるかとの問いには、「ほとんどすべての委員が目を通している」または「目を通している委員の方が多い」と答えたIRBが、A群では51.4% (36/71), B群では74.3% (26/35), C群では63.3% (19/30) だった (問18). 積極的に参加している委員の割合は、「40～60%」と答えたIRBがいずれの群でも最も多かったが、「20%未満」と答えたIRBもA群に21.7% (15/69), B

群に22.9% (8/35), C群に24.1% (7/29) あった (問19).

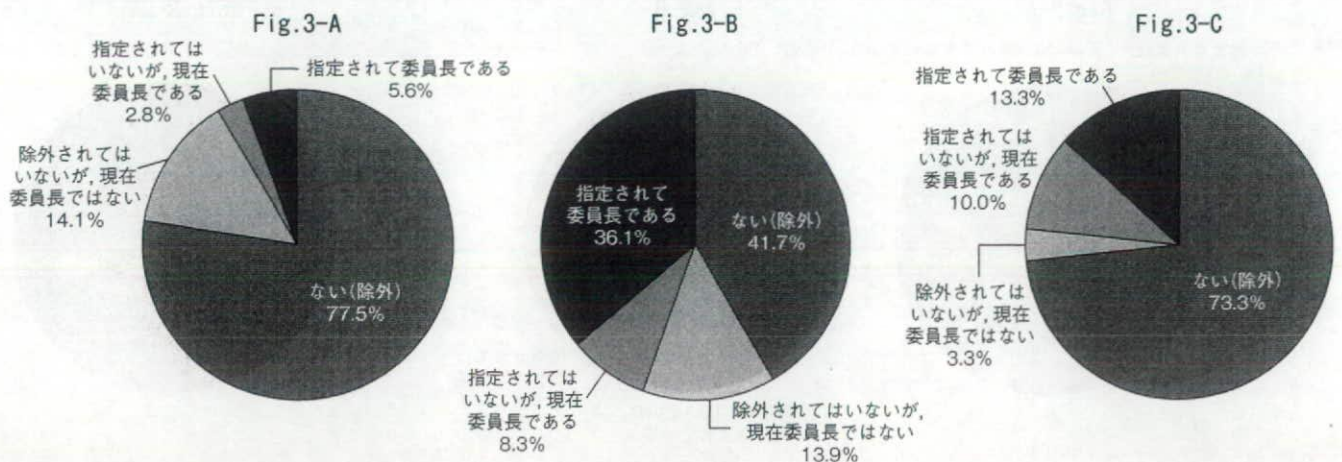
3. 委員会の審査対象について (問20～31)

まず、当該IRBが審査対象としているのはどのようなタイプの研究か、「治験」「製造販売後調査・臨床試験」「これら以外の臨床研究」から選択してもらった (問20). 「治験」と「製造販売後調査・臨床試験」を別のIRBで審査している施設はなかったため、これらは合わせて「治験、製造販売後調査・臨床試験」とし、ここで得られた結果に基づき、すでに述べたように、137のIRBをA群 (71委員会), B群 (36委員会), C群 (30委員会) に分類した. B群とC群のうち、「臨床試験 (介入研究)」を審査するIRBはB群で83.3% (30/36), C群で86.7% (26/30) を占め、「調査 (非介入研究)」を審査するIRBは、B群で72.2% (26/36), C群で53.3% (16/30) だった. 「症例報告」を審査していると回答したIRBは、B群に25.0% (9/36), C群に26.7% (8/30) あった (問21). また、ゲノム・遺伝子解析研究など、特殊な研究の審査に特化した委員会かを問うた問22への回答に基づいて、今回の解析対象からは「専門部会」を削除した.

審査件数には著しい差があり、年間100件近い新規課題を審査する委員会もあれば、調査対象の

Fig. 3 施設長が委員長になる可能性

資料の問11「施設長や委員会が設置されている部局の長 (病院長, 学部長など) が、委員長になることがありますか」への回答による. A, B, Cは、Table 2のA群, B群, C群を示す.



2005年度には1件も審査しなかった委員会が19もあった(問23, Fig. 4). 審査件数が多かったのは、概して、大学に設置されたIRBと臨床試験を専門とする医療機関に設置されたIRBであった。

研究計画の変更を審査していると答えたIRBは、「治験、製造販売後調査・臨床試験」については88.0% (88/100) だったが、「それ以外の研究」では70.3% (45/64) に留まった(問24). 継続審査については、「治験、製造販売後調査・臨床試験」では77.6% (76/98) とやはり多かったが、「それ

以外の研究」では29.7% (19/64) に過ぎなかった(問25). 重篤な有害事象の報告を行っているIRBは、「治験、製造販売後調査・臨床試験」では91.8% (90/98) だったが、「それ以外の研究」では54.7% (35/64) しかなかった(問26).

審査対象となる研究であっても申請されていない可能性がある」と答えたIRBは全体では15.6% (21/135) だった。A群では当然ながら少なく4.35% (3/69) だったが、B群では33.3% (12/36), C群では20.0% (6/30) に上った(問27, Fig. 5).

Fig. 4 審査件数

資料の問23「2005(平成17)年度の新規課題審査件数を記入してください」への回答による。

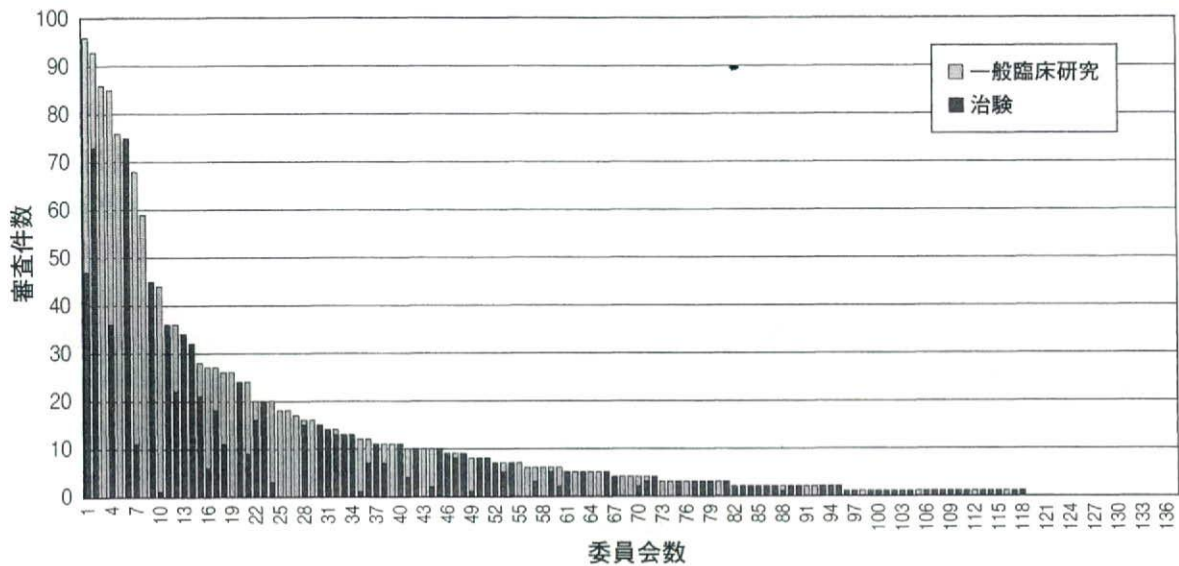
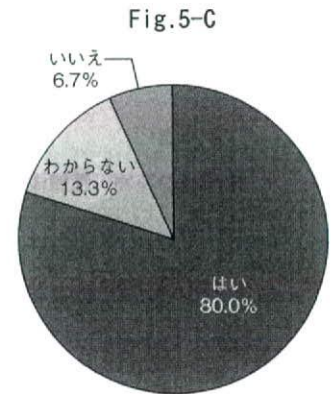
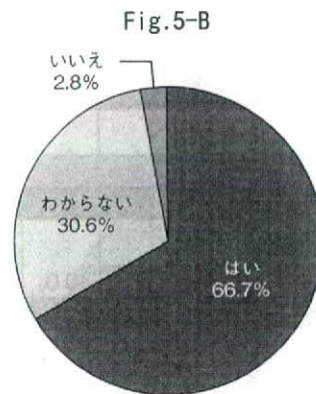
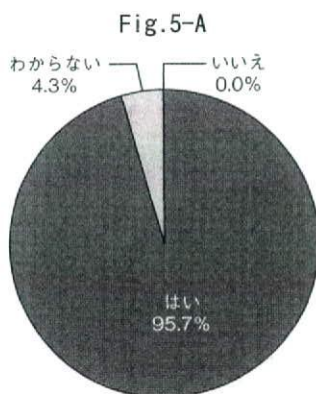


Fig. 5 審査されていない研究の有無

資料の問27「審査対象範囲内の研究は、もれなく委員会に申請されていますか」への回答による。A, B, Cは、Table 2のA群, B群, C群を示す。



*3 「研究」に対峙させるべきなのは、正確には「治療」ではなく「診療」であったが、アンケート調査票作成時にはこのことに気づいていなかった。「診療」には「治療」のほか「予防」と「診断」が含まれる。

また、審査対象となる研究であっても何らかの判断で審査は不要とされる可能性がある」と答えたIRBが全体で23.7% (32/135) もあり、A群24.6% (17/69)、B群27.8% (10/36)、C群16.7% (5/30) だった (問28)。

問29～31はやや特殊な設問だが、治療^{*3}と研究の違いをどう考えるかについて尋ねたものである。審査対象が研究である以上、研究の定義が明確でなければならないためである。治験は明らかに研究なので、研究か否かが問題になることはないと思われたため、これらの問いはB群とC群だけを対象とした。申請された行為が治療なのか研究なのか議論になったことがあると答えたのは、B群で20.0% (7/35)、C群で16.7% (5/30) だった (問29)。また、治療と研究を区別する明確な基準があると答えたのは、B群で14.3% (5/35)、C群で16.7% (5/30) に過ぎなかった (問30)。研究に当たると考える行為を選択してもらったところ、「未承認の薬物・医療機器を用いる場合」を選んだIRBが最も多く、B群で52.8% (19/36)、C群で73.3% (22/30) を占め、一方、「症例報告」を選択したのは、B群で22.2% (8/36)、C群で30.0% (9/30) に過ぎなかった (問31, Fig. 6)。「一

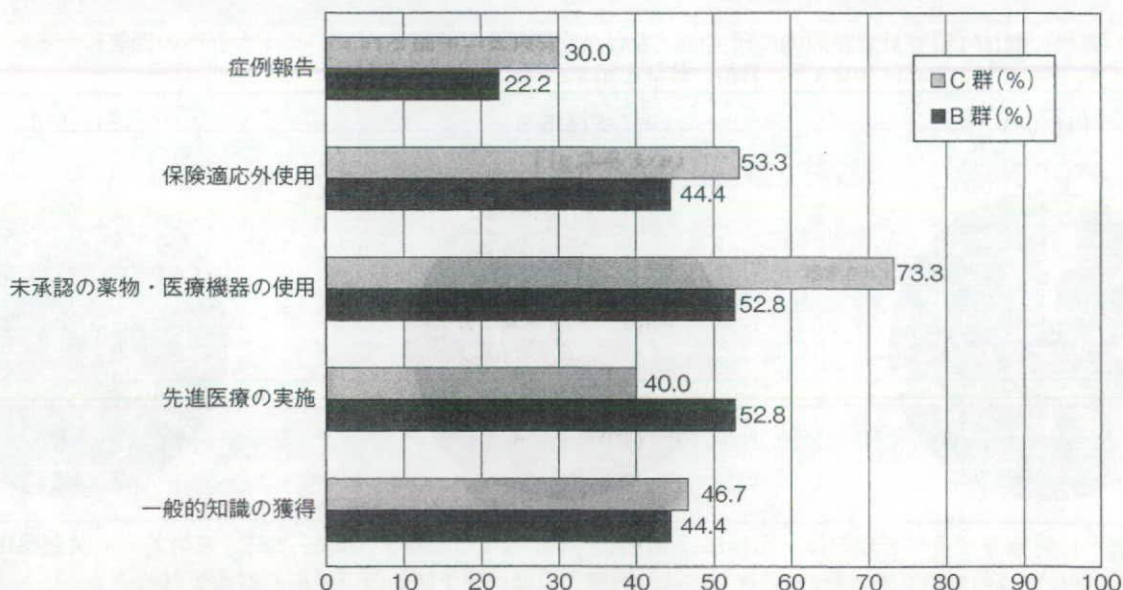
般的知識の獲得という目的が存在する場合」(「ベルモント・レポート」等)という研究の定義に該当)を選択したのは、B群で44.4% (16/36)、C群で46.7% (14/30) に留まった。

4. 委員会の実施方法について (問32～44)

「ほぼ定期的」に開かれるIRBは、A群では74.6% (53/71)、C群では72.4% (21/29) だが、B群では33.3% (12/36) に過ぎなかった (問32)。開催頻度は、A群とC群では「1ヶ月に1度程度」が最も多く、A群で52.1% (37/71)、C群で51.7% (15/29) を占めたが、B群では「半年に1回程度」が30.6% (11/36) で最も多かった (問33, Fig. 7)。本審査の前に予備調査 (事前検討) を「行っている」と答えたIRBは、A群では25.4% (18/71)、B群では27.8% (10/36) だったが、C群では48.3% (14/29) に上った (問34)。具体的には、治験の場合は事前ヒアリングを行う施設が多かったが、その他の研究については、少数ながら、専門家による予備審査を行っている施設もあった。審議課題の概要書を作成しているIRBは、A群48.6% (34/70)、B群41.7% (15/36)、C群58.6% (17/29) だった (問35)。審議資料を1週

Fig. 6 研究に該当する行為

資料の問31「以下のうち、「治療」の範囲を越えて「研究」に相当するとお考えになるものを選び下さい (複数回答可)」への回答による。



問以上前に配布しているIRBは、A群39.4% (28/71), B群47.2% (17/36), C群37.9% (11/29)で、大部分のIRBが3日以上前までには配布していた(問36)。しかし、事前配布していないIRBも、A群に16.9% (12/71), B群に13.9% (5/36), C群に6.90% (2/29)あった。1件当たりの審議時間は、いずれの群でも「30分程度」が最も多かった(問37)。委員会全体にかかる時間としては、A群では「1時間未満」が70.0% (49/70)で最も多かったのに対し、B群とC群では「1~2時間」が最も多く、B群で47.2% (17/36), C群で55.2% (16/29)を占めた。「3~4時間」と答えたIRBがあったのはC群だけで、その10.3% (3/29)を占めていた(問38, Fig. 8)。主任審査委員制をとっ

ているIRBは、全体で3.68% (5/136)と少なかった(問39)。欠席委員から事前に意見を聞いていると答えたIRBは、全体で21.5% (29/135), A群17.1% (12/70), B群27.8% (10/36), C群24.1% (7/29)だった(問40)。問41と問42は、治験(および製造販売後調査・臨床試験)に関する設問で、A群とC群のみを対象とした。依頼者の出席は、A群では「必要に応じて求めている」(38.6%, 27/70), C群では「普通は求めている」(48.3%, 14/29)と答えたIRBが最も多かった(問41)。責任医師(または分担医師)の出席は、A群もC群も「原則として求めている」が最も多く、各々58.0% (40/69), 65.5% (19/29)だった(問42)。一方、それ以外の臨床研究については、B群とC

Fig. 7 開催頻度

資料の問33「委員会の開催頻度はどの程度ですか」への回答による。A, B, Cは、Table 2のA群, B群, C群を示す。

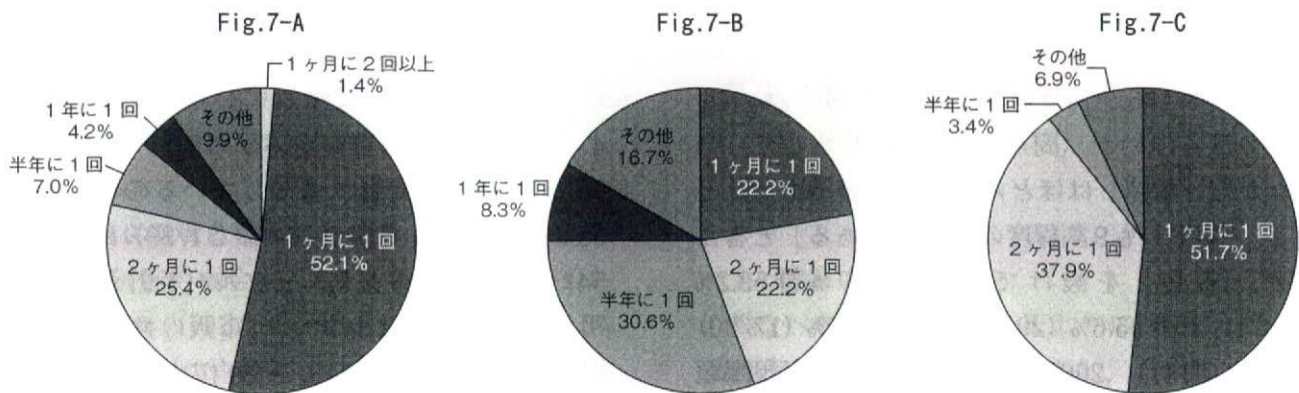
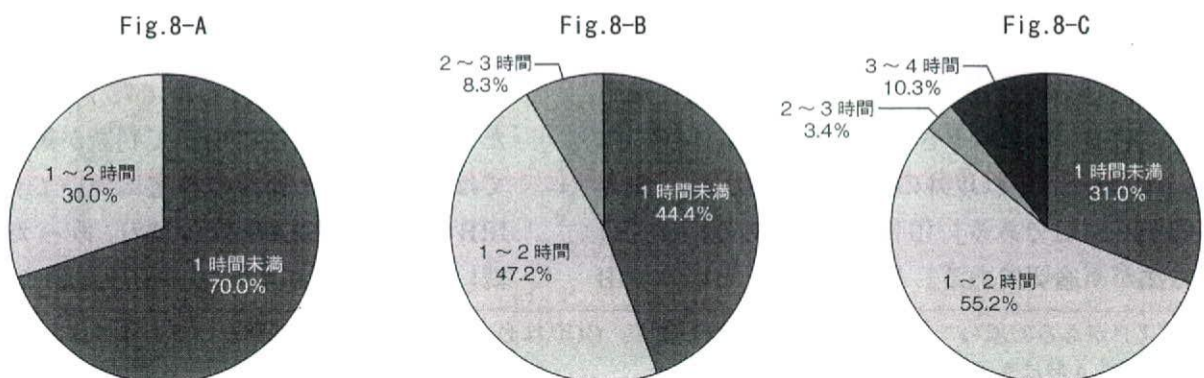


Fig. 8 かかる時間

資料の問38「委員会全体では平均するとどれくらいの時間がかかりますか」への回答による。A, B, Cは、Table 2のA群, B群, C群を示す。



群のみを対象としたが、研究責任者（または分担者）の出席を「原則として求めている」IRBがB群、C群とも多く、各々72.2% (26/36)、69.0% (20/29) だった（問43）。運営上の問題点としてもっとも多かったのは、いずれの群でも「日程調整がしばしば困難」なことで、特にB群では61.1% (22/36) のIRBがこの問題を挙げた（問44）。次いで、「審議課題が多く、委員会の負担が大きい」こと（全体で21.2%, 29/137）、「事務作業が滞りがちである」こと（全体で13.9%, 19/137）が挙げられた。

5. 採決について（問45～51）

採決に当たって研究当事者を退席させているIRBは全体で72.1% (98/136) であり、同席させているIRBは、A群では14.1% (10/71) だったが、B群で30.6% (11/36)、C群では41.4% (12/29) もあった（問45）。承認のため必要な賛成数は、「全員」と答えたIRBが全体で58.8% (80/136)、「3分の2以上」が20.6% (28/136)、「過半数」が13.2% (18/136)、「多数決」が2.94% (4/136) と、かなりばらついた（問46）。採決結果として、当然ながら「承認」はほとんどのIRBが「ある」と答え、「却下」も9割程度のIRBが「ある」と答えたが、「保留」を設けているのは、A群59.2% (42/71)、B群55.6% (20/36)、C群56.7% (17/30) だった（問47）。2005年度に審査された新規課題について、承認された件数を全IRBで合計すると、「治験、製造販売後調査・臨床試験」では820課題、「それ以外の臨床研究」では756課題であり、却下された件数は、全IRBを合わせても、「治験、製造販売後調査・臨床試験」で4課題、「それ以外の臨床研究」で9課題に過ぎなかった（問48）。「治験、製造販売後調査・臨床試験」では、「危険性が利益を上回る」ため却下されることが多く（4件中2件）（問49）、「それ以外の臨床研究」では、「研究目的が不適切である」（9件中2件）あるいは「研究方法が不適切である」（9件中2件）という理由

で却下されることが多かった（問50）。審査結果を公開しているIRBは少なく、「公開している」と答えたIRBでも、多くは「求められたら開示できる」という意味であり、自主的に一般公開しているIRBはほとんどなかった（問51）。

6. 審査基準について（問52～55）

治験（および製造販売後調査・臨床試験）以外の臨床研究は何に基づいて審査されているかB群とC群のみに訊いたところ*4、「ヘルシンキ宣言」が両群とも最も多く、B群では66.7% (24/36)、C群では66.7% (20/30) がこれを挙げた（問52）。また、B群の44.4% (16/36)、C群の40.0% (12/30) が「臨床研究に関する倫理指針」を挙げた。しかし、国際医科学団体協議会（CIOMS）の倫理指針を挙げたIRBは、B群で0% (0/36)、C群で3.3% (1/30) と極めて少なかった。「独自の審査基準を設けている」と答えたIRBがC群に2つあったが、聞き取り調査により、これらは委員会の定款であり審査基準には該当しないことがわかった。「審査基準」を委員に通知しているIRBは、B群が47.1% (16/34)、C群が64.3% (18/28) だった（問53）。また、研究者に通知しているのは、B群が35.3% (12/34)、C群が40.7% (11/30) だった（問54）。何らかの審査マニュアル（手引き書）を利用しているIRBは少なく、「市販のマニュアルを用いている」IRBはなく、「独自のマニュアルを用いている」IRBが7つあるのみであった（問55）。

7. 科学性と倫理性の審議について（問56～61）

科学性について「十分審議できている」または「おおかた審議できている」と答えたIRBは、全体で79.1% (106/134)、A群で79.7% (55/69)、B群で80.6% (29/36)、C群で75.9% (22/29) だった（問56）。科学性について判断が難しい点としては、「リスクと利益の推定について」を挙げたIRBが全体で62.8% (86/137) あった（問57）。続いて、倫理性について「十分審議できている」

*4 GCPがあると思ってA群への質問は省略したが、GCPは必ずしも倫理審査の基準というわけではないので、同じ質問をA群にもするべきだったと後から思った。

または「おおかた審議できている」と答えたIRBは、全体で85.7% (114/133), A群で82.6% (57/69), B群で82.9% (29/35), C群で96.6% (28/29) だった (問58). 科学性に関する回答と一致して、倫理性について判断が困難と感じる点として、全体として最も多く挙げられたのは「プラセボ使用の適切性について」で、全体が46.7% (64/137), A群が53.5% (38/71), B群が25.0% (9/36), C群が56.7% (17/30), 次いで「被験者にとってほとんど利益のない研究の可否について」で、全体が42.3% (58/137), A群が36.6% (26/71), B群が50.0% (18/36), C群が46.7% (14/30) だった (問59). 望ましい改善策としては、「IRBのための具体的な判断基準を設ける」を選択したIRBがいずれの群でも最も多く、全体で40.1% (55/137), A群で40.8% (29/71), B群が30.6% (11/36), C群が50.0% (15/30) であった (問60). 多施設共同研究の審査を段階的に行うとして、「中心的な」IRBにかけるのは「施設」IRBの「事前がよい」とする回答が多かったが、A群では「事後」や「両方」を選択したIRBもあった (問61).

8. 委員の教育と研修について (問62～68)

教育・研修の機会が「ない」と答えたIRBは、全体で89.6% (121/135) に上り、A群では91.5% (65/71), B群では91.4% (32/35), C群では

82.8% (24/29) と、いずれの群でも圧倒的多数を占めた (問62, Fig. 9). 一方、教育・研修の機会が「ぜひあった方がよい」または「あった方がよい」と答えたIRBは、全体で80.7% (109/135) に上り、A群では80.3% (57/71), B群では77.1% (27/35), C群では86.2% (25/29) と、いずれの群でも大部分を占めた (問63, Fig. 10). 研修の機会を設ける頻度は、いずれの群も「1年に1回程度」が多かった (問64). 望ましい研修方法として多く挙げられたのは、「講義」「事例検討会」「模擬IRB」などであった (問65).

また、IRB間の意見交換の機会が「あった方がよい」と答えたIRBが全体で60.4% (81/134) を占め (問66), その方法としては「連絡協議会の設置」「ウェブの利用」「電子メールの利用」などが多かった (問67).

IRB委員の資格認定制については、「ぜひあった方がよい」と答えたIRBは全体で2.2% (3/135) に過ぎず、「あった方がよい」と答えた20.0% (27/135) を加えても、積極的に望んでいるのは2割程度と考えられた (問68, Fig. 11).

考 察

1. IRBの数は適切か

問3への回答からわかるように、IRBの数は最近急激に増加している。しかし、日本にはIRBの

Fig. 9 教育の機会

資料の問62「IRBについて委員が研修する(教育を受ける)機会がありますか」への回答による。A, B, Cは、Table 2のA群, B群, C群を示す。

