

調査時期	－4 週 以 内	0 週	4 週	8 週	12 週	18 週	24 週	28 週	36 週	48 週	72 週
A 群	観察期間		ABC a mg 1日1回					追跡			
B 群			ABC プラセボ 1日1回					調査期間			
同意	◎										
患者背景	◎										
ABC 剤/プラセボ の服薬状況			◎	◎	◎	◎	◎				
基礎治療遵守状 況			◎	◎	◎	◎	◎				
HCV ウイルス検査	◎		◎	◎	◎	◎	◎	◎	◎	◎	◎
肝機能検査	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎	◎
末梢血液検査	◎	◎	◎	◎	◎	◎	◎	○	○	◎	◎
血清電解質検査	○	◎	◎	◎	◎	◎	◎	○	○	◎	◎
血圧測定		◎	◎	◎	◎	◎	◎	○	○	◎	◎
腎機能・尿検査		◎	◎	◎	◎	◎	◎	○	○	◎	◎
自覚症状		◎	◎	◎	◎	◎	◎	◎	◎	◎	◎

◎: 必須、○; 可能な限り

## 9. 中止基準

個々の症例が以下のいずれかの中止基準に該当する場合、その症例の併用治療を中止する。

中止の日付・時期、中止の理由、経過をカルテならびに CRF に明記するとともに、中止時点で必要な検査（調査スケジュール表の 24 週時点の◎の検査）を行い有効性・安全性の評価を行う。有害事象発生により中止した場合は、可能なかぎり原状に回復するまでフォローする。また、ABC 割付センター宛に中止報告書を送付する。ABC 割付センターは各施設の試験責任医師に中止報告書を送付する。中止後も可能なかぎり 72 週までの規定の時点での観察・検査を行う。

治療に反応がないこと（効果不十分）により中止した場合、後治療を行った場合には後治療開始日・内容についてカルテならびに CRF に明記する。

### 中止基準

- 1) 被験者から試験参加の辞退の申し出や同意の撤回があった場合
- 2) 有害事象により試験治療の継続が困難な場合
- 3) 治療に反応がない（効果不十分）場合（基準の詳細は別紙に記載）
- 4) 原疾患の悪化のため、試験治療の継続が好ましくないと判断された場合

- 5) 合併症の増悪により試験治療の継続が困難な場合
- 6) 登録後に適格性を満足しないことが判明した場合
- 7) 妊娠が判明した場合
- 8) 試験全体が中止された場合
- 9) その他の理由により、医師が試験を中止することが適当と判断した場合

**【設定根拠】**

1): GCP に従う、 2)~5)、8): 倫理的観点から、6)、7)、9): 安全性確保のため

**1 0. 有害事象発生時の取扱****(1) 有害事象発生時の被験者への対応**

試験責任医師または試験分担医師は、有害事象を認めたときは、直ちに適切な処置を行うとともに、カルテならびに症例報告書に記載する。また、試験薬の投与を中止した場合や、有害事象に対する治療が必要となった場合には、被験者にその旨を伝える。

重篤又は重大な有害事象が発生し、割付け薬剤が ABC であるか否かにより対処法が異なる場合（緊急事態が発生した場合）は研究代表者にその旨を連絡し、研究代表者は割付けコード化担当者に当該被験者番号と担当医師名を伝える。割付けコード化担当者は当該症例についてのみ割付けコードを開封し、当該症例の担当医師に割付けられた治療法を伝える。

**(2) 重篤な有害事象の報告**

重篤な有害事象とは以下のいずれかの定義に該当する好ましくない事象とする（薬事法施行規則第 253,273 に準じる）。

- 1) 死亡または死亡につながるおそれ
- 2) 入院または入院期間の延長
- 3) 障害または障害につながるおそれ
- 4) 後世代の先天性の疾病または異常

報告の対象となる有害事象は、試験期間中の全ての重篤な有害事象、試験治療終了（中止）後に試験薬との関連性が疑われる重篤な有害事象とする。

試験責任医師は、重篤な有害事象の発生を認めたときは、速やかに病院長（臨床試験部？）に報告する。本試験は多施設共同試験であるため、他の実施機関の長にも通知する。報告は第一報（緊急報告）および第二報（詳細報告）とする。

報告の期限については、薬事法施行規則第 253,273 に従い、7 日以内（未知、死亡または死亡につながるおそれがある）および 15 日以内（未知、重篤）とする。施設ごとの様式・規定があれば、それにも従う。

また、ABC は市販後の薬剤であるため、厚生労働省の医薬品等安全性情報報告制度により（独）医薬品医療機器総合機構（ABC 剤の製造販売会社・・・を通して）に報告する。

### (3) 重要な有害事象の報告

重篤な有害事象の基準を満たさないが、重要な有害事象の条件を満たす場合は、速やかに重篤な有害事象の報告に準じて報告を行う（薬事法施行規則第 253,273 に準じる）。

- ・感染症、がん など

### (4) その他の有害事象

その他の有害事象については、「8. 観察および検査項目 (11) 有害事象と副作用の確認」に記載した手順により、症例報告書に記載する。

## 1 1. 実施計画書からの逸脱の報告

試験責任医師または試験分担医師は、研究代表者の事前の合意および治験審査委員会の事前の審査に基づく病院長の承認を得る前に、試験実施計画書からの逸脱あるいは変更を行ってはならない。

試験責任医師または試験分担医師は、緊急回避等のやむを得ない理由により、研究代表者との事前の合意および治験審査委員会の事前の承認を得る前に、試験実施計画書からの逸脱あるいは変更を行うことができる。その際には、試験責任医師または試験分担医師は、逸脱または変更の内容および理由ならびに試験実施計画書等の改訂が必要であればその案を速やかに、研究代表者および治験審査委員会に提出し、研究代表者、治験審査委員会および病院長の承認を得るものとする。

試験責任医師または試験分担医師は、試験実施計画書からの逸脱があった場合は、逸脱事項をその理由とともに全て記録し、試験責任医師は、医療機関で定めた所定の様式により病院長および研究代表者に報告しなければならない。試験責任医師は、これらの写しを保存しなければならない。（GCP 第 46 条に準じる）

## 1 2. 試験の終了、中止、中断

### (1) 試験の終了

各施設での試験の終了時には、試験責任医師は、速やかに試験終了報告書を病院長及び研究代表者に提出する。

### (2) 試験の中止、中断

IRB により、中止の勧告あるいは指示があった場合は、試験を中止する。試験責任医師（研究代表者を含む）で構成する委員会は、以下の事項に該当する場合は試験実施継続の可否を検討する。

- 1) 試験薬の品質、安全性、有効性に関する重大な情報が得られたとき。
- 2) IRB により、実施計画等の変更の指示があり、これを受入れることが困難と判断されたとき。

試験の中止または中断を決定した時は、速やかに病院長（あるいは各医療機関の長）にその理由とともに文書で報告する。試験中止の決定を行った後、関連施設治験審査委員会等、関係医師、薬剤の提供を受けた会社関係者等に速やかに伝達し、中止後の処理にあたるものとする。

### 1 3. 実施計画書等の変更

実施計画書や同意説明文書の変更（改訂）を行う場合は予め各医療機関の治験審査委員会あるいは倫理委員会の承認を必要とする。変更の決定は関連施設の長、関係医師等に速やかに伝達する。また、補足、解釈などの細目の変更については、研究代表者が、必要な場合には適切な関係者に意見を求めた上で決定し、その結果を関係医師等に伝達し、治験審査委員会あるいは倫理委員会に報告し了承を得る。

### 1 4. 試験実施期間

被験者の登録期間：2009 年 3 月～2009 年 11 月

試験実施期間：2009 年 3 月～2010 年 11 月

ただし、目標症例数に達し次第、被験者の登録を終了する。

### 1 5. 統計解析

#### (1) 有効性および安全性の解析の対象集団

有効性および安全性の主要な解析対象は、無作為割り付けをされた全ての被験者から ABC 薬または ABC のプラセボ未投与例を除いた全ての被験者である最大の解析対象集団（FAS; Full Analysis Set）とする。副次的な解析対象集団として試験実施計画書に適合した対象集団（PPS; Per Protocol Set）とする。なお、試験実施計画書に適合した集団は以下に定義する適合格例の全例とする。データの取り扱いは評価項目ごとに定める。

#### (2) 症例の分類の定義

適合格例：選択基準のすべてを満たし、除外基準のいずれにも該当しない症例

中止症例：9 章の中止基準により試験を中止した症例

#### (3) 中止症例、欠測値などのデータの取扱い

##### 1) 中止症例のデータの取扱い

- ・ FAS については、中止時期によらず中止時点まで実際に測定された値を評価に用いる。
- ・ PPS については、中止時期が治療開始から 4 週未満である場合は、欠測値として取扱う。中止時期が 4 週以降である場合は、実際の測定値を用いる。

##### 2) HCV RNA の欠測値の補填

検定の際は、試験治療中の規定の評価時点については、規定の評価時点に最も近い前（または後【どちらかに規定しておく】）の評価を、規定の評価時点が併用投与終了後の場合は評価時点に最も近い後（又は前【どちらかに規定しておく】）の評価を代入する。【つまり、「欠測が生じた時点まで

の測定値のうち、最後に測定された値で欠測値を代用し、評価を行う（LOCF; Last Observation Carried Forward）」など、各自のテーマに適切な取り決めをしておくこと】

中止症例について、(3) 1)の取り扱いで PPS の欠測値とされた場合には、非陰性化（無効）として扱う。

次の治療法違反後の観測値は FAS 及び PPS に対しそれぞれ以下のように定める。

ABC 又はプラセボの服薬不遵守は FAS に対し・・・PPS に対し・・・、併用禁止薬の使用は FAS に対し・・・PPS に対し・・・、基礎治療不遵守は FAS に対し・・・PPS に対し・・・。

【「・・・」に各自のテーマに適切な取り扱い方法を検討すること】

規定の評価時点ごとのデータの要約では、規定の評価時点の許容範囲を超える観測日のズレがある場合や、既定以外の方法又は条件により得られた測定値は欠測値（対象から除外）として取り扱う方法も参考までに行う。

### 3) HCV RNA 以外の欠測値

検定の際は HCV RNA の欠測値の代入方法に従う。

投与前値の欠測値は・・・

上記に定めていない症例やデータの取り扱いは、研究代表者、試験統計家及び試験責任医師で協議、決定し、統計解析のための解析対象の採用・不採用を含めたデータの固定を行う。開錠はデータの固定後に行う。

## (4) 解析方法

### 4.1) 症例の内訳

FAS、PPS、中止症例数を群の識別とともに表示する。

### 4.2) 治療の状況

ABC 剤又はプラセボの服薬遵守状況(規定の投与回数に対する実施回数の比率)について集計する。

### 4.3) データの要約

連続値として得られる全ての検査項目について、治療法ごと検査時点ごとに基礎統計量（最大値、中央値、最小値、25%点、75%点、平均値、標準偏差）を算出する。HCV RNA の陰性化率を、治療方法ごとにそれぞれの時点で算出する。肝機能検査（AST(GOT)、ALT(GPT)、血清ビリルビン、ALP、 $\gamma$ -GTP、ZTT、血清蛋白、血清アルブミン、総コレステロール）は投与開始前からの差、および投与開始前に対する比の対数変換後のデータについても基礎統計量を算出し、個々の被験者の経時推移を図（折れ線グラフ）に示し、経時推移の要約を Box-Whisker plot に示す。そのほかの経時的に測定されたデータは、経時推移の要約を Box-Whisker plot に示す。

カテゴリーデータとして得られる全ての検査項目について、治療法ごと検査時点ごとにカテゴリーの集計を行う。

#### 4.4) 背景因子の解析

年齢、体重、AST(GOT)、ALT(GPT)、についてデータの分布を 4.3)に従って要約する。性別、HCV RNA の多少、HCV RNA のジェノタイプ、合併症（有り無し）、飲酒歴（有り無し）について治療法ごとにカテゴリーの集計を行う。

#### 4.5) 主要評価項目の解析

- 1) 評価項目：HCV RNA の陰性化率
- 2) 評価時点：投与終了後 24 週
- 3) 評価方法：

陰性化率の比較に Fisher の正確な検定を適用する。有意水準は両側 5%とする。また、陰性化率の差の 95%信頼区間を正規近似を用いて計算する。

副次的に、6. 試験の方法（6）割り付け方法で指定した層別因子の調整を行う Cochran-Mantel-Haenszel 検定を適用する。

#### 4.6) その他の主な評価項目の解析

- 1) HCV RNA の投与終了後 48 週の陰性化率

Fisher の正確な検定を適用する。有意水準は両側 5%とする。また、陰性化率の差の 95%信頼区間を正規近似を用いて計算する。

副次的に、6. 試験の方法（6）割り付け方法で指定した層別因子の調整を行う Cochran-Mantel-Haenszel 検定を適用する。

- 2) 肝機能検査項目

ALT(GPT)の投与終了後 24 週および 48 週の正常化率について時点ごとに Fisher の正確な検定を適用する。有意水準は両側 5%とする。また、正常化率の差の 95%信頼区間を正規近似を用いて計算する。

副次的に、6. 試験の方法（6）割り付け方法で指定した層別因子の調整を行う Cochran-Mantel-Haenszel 検定を適用する。

- 3) 有害事象

有害事象の一覧を作成するとともに、有害事象ごとの発生率の比較に Fisher の正確な検定を適用する。検定は検査項目ごとに両側検定、有意水準 5%とする。

#### 4.7) その他の評価項目の解析

末梢血液検査、血清電解質検査の経時推移については投与開始時点の測定値からの差の要約を Box-Whisker plot に示すとともに、投与開始後 24 週時点の投与開始時点の測定値からの差の基礎

統計量を算出し、Wilcoxon の順位和検定を適用する。

## 16. 目標症例数および設定根拠

### (1) 目標症例数 これから計算

A 群・・・例、B 群・・・例、計・・・例

1 施設目標症例数：\*\*例（A 群\*例、B 群\*例）、参加施設数\*\*

### (2) 設定根拠

報告された試験成績より ABC a mg 併用群の陰性化率（投与終了後 24 週時の HCV RNA 検査が検出限界未満である患者の割合）を 90%、ABC のプラセボ併用群の有効率を 70%と推定した場合、両側検定（有意水準 5%）で検出力 80%を保持するのに必要な標本の大きさは各群・・・例、合計・・・例必要となる。

【注意；ここでは、練習のために症例数を算出するための式も記載すること。】

## Randomized Controlled Trials: Design and Statistical Analysis

### 臨床試験概論I

Toshiro TANGO

Department of Technology Assessment and Biostatistics  
National Institute of Public Health, Japan

October 27, 2008

## Outline

- What is a RCT ?
  - Historical background
  - Impact of RCTs
  - Importance of Statistical Inference
  - Evidence and belief
  - Poor environment in Japan
  - Ethical issues
- Structure and Justification of RCTs
  - Eligible patients
  - The need for concurrent controls
  - The importance of randomization
- Bias in RCTs
  - Selection bias
  - Allocation bias
  - Assessment bias
  - Placebo-effects
- Statistical Analysis
  - When to stop
  - Ethical issue
  - The primary endpoint
  - The adequate evidence
  - Hypothesis tests
  - Sample size
  - Summary statistic of primary endpoint

## What is a randomized controlled trial ?

- Simply an experiment, performed on human subjects to assess the efficacy of a new treatment for some condition.
- The new treatment is given to a group of patients and another treatment is to another group of patients at the same time; "concurrently controlled".
- Patients are allocated to one or other group by **randomization**, something like "the toss of a coin". For this reason, there is an important ethical issue to the conduct of RCTs, which sets them apart from other experiments such as agriculture or industry.

## What is a randomized controlled trial ?

### Historical background

In the past, the effectiveness of treatments has often been decided by reference to **ad hoc** usage in the hands of some eminent authority. Indeed, this approach was widespread until quite recently. RCTs in forms that would currently be recognized made their entrance only in the period since the second World War. However, since medieval times, there have been isolated attempts to obtain empirical evidence on the effectiveness of treatments.

## What is a randomized controlled trial ?

### Impact of RCTs

- The RCT is an introduction of the scientific method into the process of comparing treatments.
  - 1) Our test of whether one treatment is better than another is based upon **observing the treatments** when they are applied to patients.
  - 2) It is **not** based on any **theory** of how the treatments might work.
  - 3) It is **not** based on **anecdotal** evidence, perhaps gained from a doctor 'trying out' the treatment in an uncontrolled manner.
  - 4) It is **not** based on any appeal to **authority**.

## What is a randomized controlled trial ?

### Impact of RCTs (II)

Over the past 50 years RCTs have become established as the primary and, in many instances, the only acceptable source of evidence for the efficacy of new treatments. The points (3) and (4) above, which used to be commonplace, may even be illegal. The use of empirical evidence, not anecdotal, to settle the matter is what gives statistics its great importance in RCTs.



## What is a randomized controlled trial ?

### Importance of Statistical Inference

- Not all patients will react to treatment in the same way !
- So, whether the *difference between* the two treatment groups is significant, given the *variation* in the response observed *within* each group, is clearly a statistical question. →
- test statistic = between-difference / within-variation**
- Furthermore, the RCT is of little value unless the results obtained can be generalized to as yet untreated patients; this *inference* from the *sample* to the *population* is a statistical exercise.

## What is a randomized controlled trial ?

### Importance of Statistical Inference (II)

- As with any inferential process, the *quality* of the inference will be higher if the data used are *collected appropriately*.
- Again, statistical ideas are of the utmost importance in the design of RCTs.

## What is a randomized controlled trial ?

### Evidence and belief

- Therefore, RCTs are conducted to gather empirical evidence about *differences between treatments*.
- However, there is a delicate balance between *evidence* and *belief* in this area.
  - It is unlikely that a treatment will ever be subjected to the rigors of an RCT if no one believes it shows an improvement on the status quo.
  - Therefore, people involved in the conduct of RCTs *believe...new treatment may be superior*.
  - However, these people will also acknowledge that their beliefs are essentially unsupported and might even be mistaken.

## What is a randomized controlled trial ?

### Evidence and belief (II)

- Simply believing that one treatment is superior to another is not a justification for acting on that belief.
- Such justification requires you to collect evidence to prove or refute your beliefs and
- The RCT is the currently accepted tool for doing this.
- Therefore, RCTs are of the utmost importance to modern medicine and statisticians are of the utmost importance to RCTs

## What is a randomized controlled trial ?

### Poor environment in Japan

RCT = Physicians + Statisticians + others  
(design + conduct) (design + evaluate)

- ... however, ...
- Quite many physicians do not understand what RCT is !
  - Only a few statisticians have the ability of making sound advice for the design and analysis of RCT.

## What is a randomized controlled trial ?

### Ethical Issues

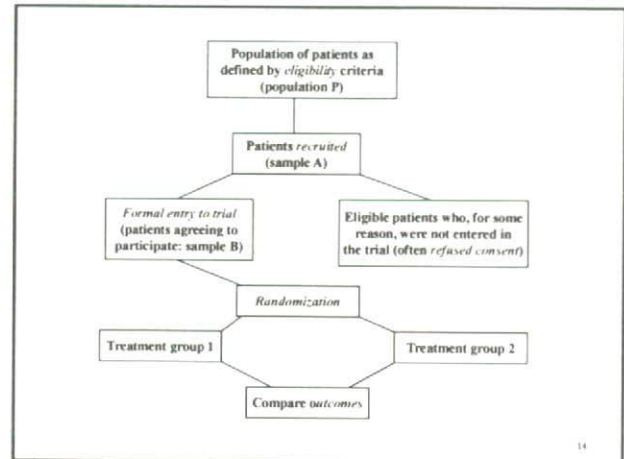
- Due to the important feature of RCTs that the **experimental units are patients**, there are various ethical responsibilities on the investigator. The following three issues are of the utmost importance.
- (1) A patient must never be given a treatment that is known to be inferior !
- (2) Patients must be fully informed about all the circumstances surrounding the treatments in the trial, including possible adverse reactions and side-effects they may experience.
- (3) Patients who have entered a trial may withdraw at any time and they must then receive the most appropriate treatment outside the trial.

## Structure and justification of RCTs

The process of RCTs possesses five key items:

- (1) a population of *eligible* patients
- (2) a group of patients *recruited* from the population
- (3) existence of (at least) *two* treatment groups
- (4) allocation to treatment is by *randomization*
- (5) *outcome measures* in the treatment groups are compared at the end of the trial

13



14

## Structure and justification of RCTs

### Eligible patients

The primary aim of the trial is to *generalize its results to all patients* who are like those treated in the trial. So, the eligibility criteria is needed to *define clearly* the patients who might enter the trial.

Example : (SNMC 26-weeks for chronic hepatitis C)

- *Inclusion criteria*
  - anti HCV and HCV-RNA *positive*
  - elevated ALT ( $> 2 \times$  ULN)
  - not eligible for interferon-ribavirin therapy
- *Exclusion criteria*
  - other acquired or inherited causes of liver diseases
  - complication of liver cirrhosis / decompensation

15

## Structure and justification of RCTs

### The need for concurrent controls

... If you want to see if a new treatment 'works', why not give it to a patient and see if he or she gets better?

Bradford Hill, writing of the pre-streptomycin era, pointed out that tuberculous meningitis was always fatal, so the recovery of any patient treated with streptomycin provided evidence of the value of the drug.

However, instances where the outcome of a disease is invariable are, these days, extremely rare and can be ignored. Instead, patients, their disease and their reaction to treatment are *all variable* to some extent.

Measuring and accounting for *this variability* is what gives *statistics* its central role in RCTs.

16

## Structure and justification of RCTs

### The need for concurrent controls (II)

- If recovery is uncertain (*within-group variation*), and we give all patients the new treatment, then we cannot differentiate *which* of the patients who recover *do so because of the treatment* or of the other factors.
- The solution is to include a second group of patients, often called *concurrent control group* or simply *control group*, in the trial who do not receive the new treatment.
  - 1) Some patients in the treatment group may have recovered anyway, but
  - 2) there will be a similar subgroup in the control group and their effect *cancels out* in the comparison.

17

## Structure and justification of RCTs

### The need for concurrent controls (III)

The control group may receive a treatment, it is just not the new treatment. In fact, if there is an effective treatment, then it would be unethical to withhold this from the control group.

The trial is then a comparison of the new treatment with the *best currently available treatment*.

If there is no accepted treatment available, the control group may receive *no treatment at all* (that is, what they would usually have received), or

They may receive a sham treatment known as a *placebo*. The use of placebo raises a number of important issues (*to be discussed later*).

18

## Structure and justification of RCTs

### The need for concurrent controls (IV)

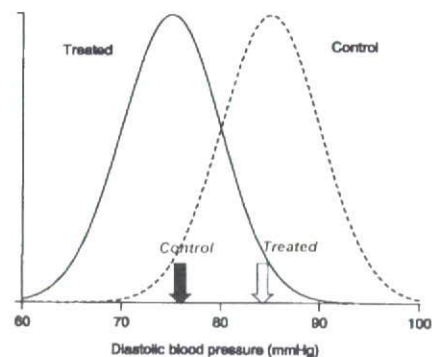
The treatment groups are usually assessed on the basis of *suited summary measures (primary endpoint)*, such as their *means*.

That is, we try to examine the data to see if there is evidence that *the mean of the treated group differs from that of the control group*.

We do not look at the effectiveness on a patient-by-patient basis, so there may well be instances where *some patients do worse on the 'superior' treatment than some patients on the 'inferior' treatment*.

Comparison of the treatments is therefore *a comparison of the distributions of the outcome measures, which definitely needs statistical methods*.

19



20

## Structure and justification of RCTs

### The importance of randomization

There are three primary justifications for the use of random allocation:

- 1) In principle at least, it ensures that the group to be given the different treatments are *comparable*.
- 2) The allocation to treatment is *unknown at the time of entry to the trial*. → *blind, masking*
- 3) There is a *statistically sound estimates of error (within-group variation)* available for the comparison of the two treatment groups.

21

## Structure and justification of RCTs

### The importance of randomization (II)

- The purpose of an RCT is to assess the relative merits of the treatments being studied. If the treatment groups differ at the end of the trial, then it is important that *we should be able to assert that this is because of the different treatments*.

If the treatment groups differed systematically before the trial ever started, then *this is impossible!!*

Randomization is the means by which the ability to state *"the difference in treatment groups is caused by the difference in treatments"* is achieved.

22

## Structure and justification of RCTs

### The importance of randomization (III)

In randomized groups,

- The proportion of *female* patients should be similar
- The proportion of patients with *blue eyes* should be similar
- The distribution of *severity of disease* should be similar
- The range of *prognostic factors (patient's features related to the outcome)* should be similar → *Confounding factors* and so forth...

An important property of randomization is that it should achieve *comparability with respect to all prognostic factors, including those that are unknown to the investigators*.

23

## Bias in RCTs

At the conclusion of an RCT, the investigators are eager to calculate a *number* that measure the difference between the treatments.

The presence in this number of some degree of *sampling error (between-patients error)* is inescapable,

But *excluding* from this number *any systematic effect* other than the treatment effect is the *overriding aim* of much of the methodologies that surrounds RCTs.

Such unwanted systematic effects are called *biases*.

24

### Bias in RCTs

#### Types of bias

In the course of conducting an RCT, many practical problems that could lead to bias will be encountered. It is useful to describe some of these ...

- (1) Selection bias
- (2) Allocation bias
- (3) Assessment bias
- (4) Publication bias

23

### Bias in RCTs

#### Selection bias

Selection bias can occur when the decision to enter a patient into an RCT is *influenced by knowledge* of which treatment the patient will receive when entered. For example,

*A doctor is thinking ... submitting a patient with severe disease to the placebo group is no good!*

this kind of selection can occur in non-randomized trials.

- In a well-run RCT, selection bias should not occur, because the patient will be formally entered into the trial before the treatment is chosen (*Formal entry procedure*).

Once a patient is formally entered into the trial, they are entered in a *patient log*, and *randomized*. They cannot then be removed from the group allocated.

24

### Bias in RCTs

#### Allocation bias

Patients have many factors that can affect the outcome of their therapy, regardless of the treatment group they are allocated. These *prognostic factors* might be

- whether tumor is advanced or not
- Whether initial disease control was good or not
- Whether or not patient can walk
- Higher or lower natural levels of immunity, ...

Randomization will, in principle, lead to groups that are balanced, in the sense that the distribution of these prognostic factors will be similar.

- However, *simple randomization*, by the toss of a coin, is obviously a stochastic phenomenon and there may be particular trials where *balance on an important prognostic factor may not be achieved*

27

### Bias in RCTs

#### Allocation bias (II)

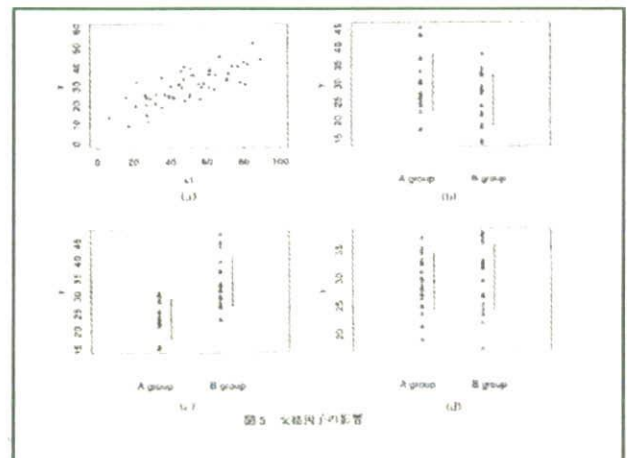
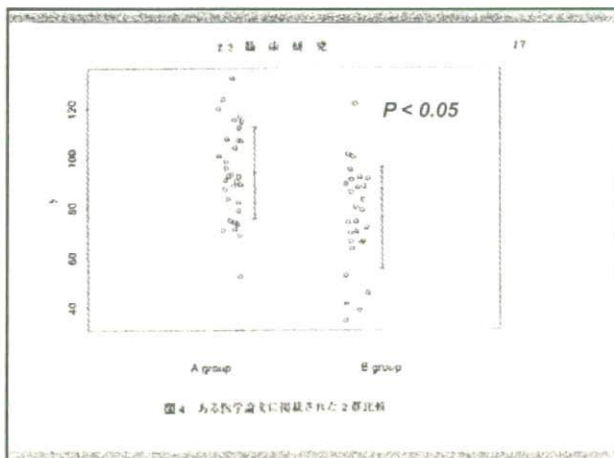
For example, consider a trial of therapies for *multiple sclerosis* in which one treatment group has a *much higher proportion of non-ambulatory patients* than the other.

The treatment in the group with a higher proportion of non-ambulatory patients is *very likely to compare poorly* with other treatments in the trial. *Why?* simply because ... *the patients in this group have more severe disease.*

The failure to form comparable treatment groups means that comparison of treatments is *biased* and we refer to this form of bias as *allocation bias*.

*One solution to this bias is to adopt a more sophisticated randomization method such as stratification method and minimization method*

28



## Bias in RCTs

### Allocation bias (IV)

Randomization is not a complete tool to adjust for the confounding factors. The remaining bias due to imbalance of prognostic factors can be adjusted for by a statistical analysis

Outcomes	Statistical Methods
Continuous	Analysis of Covariance
Binary	Logistic regression model
Time to event	Cox proportional hazard model

11

## Bias in RCTs

### Assessment bias

- At the end of the trial, and often during its course, observations are made on a variety of outcome variables.
- Many of these will be entirely *objective*, such as ALT, Hb, HCV-RNA, ... Other variables are *less objective*, such as measure of quality of life, symptom score, ...
- If the *observer knows the treatment* being given to the patient and if the measurement of an outcome variable contains an *element of subjectivity*, then it is possible that the *value of an observation might be influenced by the knowledge of the treatment*. This is called "assessment bias".

12

## Bias in RCTs

### Assessment bias (II)

- A potentially *important source of bias* in a trial arises when *either the patient or the doctor knows* which treatment the patient is receiving because many involved in trials are *hopeful that the new treatment will be better!*
- For this reason, many trials are run as single- or double-blind trials.
  - The importance of making a trial blind depends on the *objectivity of the outcome(s)* of the trial.
  - Some study simply cannot be blind. An example will be trials comparing surgical with non-surgical treatments.
  - In some studies, the statistical analysis is also done blind.
- It should be noted that blindness is used to exclude assessment bias and also to *exclude the suspicion* of assessment bias.

13

## Bias in RCTs

### Placebos

- Placebos are treatments that look like the real thing but contains no active ingredient.
  - Clearly there are no such things as "*placebo operations*" although there were trials using "sham" surgery.
  - We sometimes give *placebo injections* but
  - Major use of placebos is in trials where the treatment is an orally administered preparation, usually a *tablet or capsule*.
- Placebos, needless to say, must look, smell, feel and taste the same as their active counterparts.
- Placebos have essentially *two roles*: one is to take account of the *placebo effect* and the other is to achieve *blindness*.

14

## Bias in RCTs

### Placebo-effect

- When no standard therapy is available, investigators have to decide how to handle the control group. One simple and practical solution is to *give patients "no treatment at all"!*
- However, a problem arises because of what is known as the *placebo effect*. This is when a patient exhibits a response being given a treatment, even though the treatment has no active component.
- In other words, this can arise simply because *the patient reacts positively to "something being done"* about their condition.

15

## Bias in RCTs

### Placebo-effect

- Let us consider the following three groups
 

A: New treatment group	=	New treatment effect + placebo effect
U: untreated group	=	Nothing
P: placebo group	=	placebo effect
- Then, if we compare A and U, we cannot estimate the effect of new treatment since
 
$$\text{Group A} - \text{Group U} = \text{New treatment effect} + \text{placebo effect}$$
- However, if we compare A and P, we can estimate that effect since the placebo effect cancel out :
 
$$\text{Group A} - \text{Group P} = \text{New treatment effect}$$

16

## Statistical analysis

### When to stop ?

An RCT is used to gather "adequate evidence" about the relative merits of two or more treatments. In order to do this, we recruit eligible patients, allocate them randomly to the treatments under study and then assess their outcomes. One problem is

*How do you know when to stop ?*

- 1) you keep recruiting until the question of which treatment is superior is settled.
- 2) you *decide beforehand* how many patients you need to recruit and you only analyze when this number has been recruited. Superficially, alternative 1 seems much the more attractive and so-called *Group sequential designs* are needed. However, its implementation is not so easy. Alternative 2 is, in fact, the approach most often used in practice.

37

## Statistical analysis

### When to stop ?

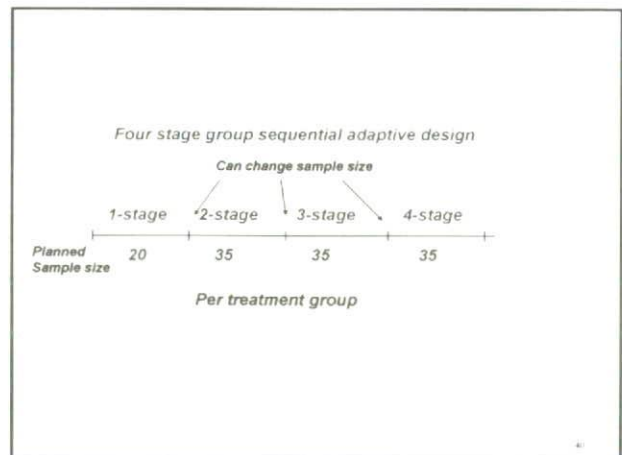
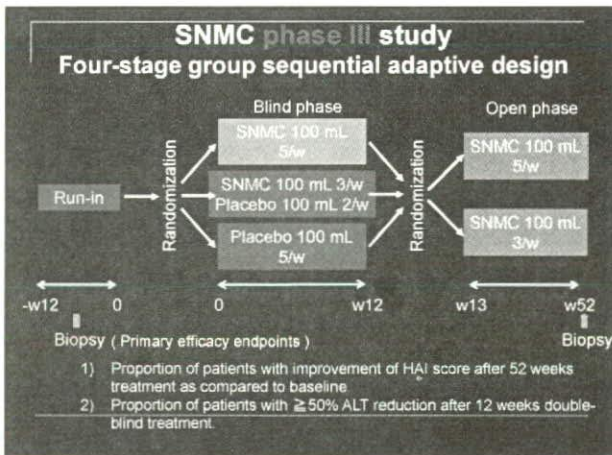
An Example of RCT with Group sequential designs is "A 52 week multicenter, randomized, double-blind placebo-controlled trial evaluating the efficacy and safety of SNMC in patients with chronic hepatitis C" which has been and is being conducted in Europe.

This study is conducted using a *four-stage group sequential adaptive design* with *sample size adjustments* after the planned interim analysis.

Primary endpoints:

- 1) Improvement of HAI score after 52-week treatments
- 2) >50% ALT reduction after 12-week double blind treatment

38



40

## Statistical analysis (II)

### Ethical issue

Obviously, the more patients you recruit the more evidence you obtain. However, you need to specify a total number of patients that is, in some sense, *adequate for your purposes*.

- 1) If you recruit *too few patients* then you may be unable to settle questions that are central to the aim of the RCT. This is not only scientifically inadequate, it is *ethically unsound*.
- 2) On the other hand, it is *also unethical* to recruit *many more patients* that you need to settle the primary question asked in the RCT.

*"patients allocated to inferior treatment"*

41

## Statistical analysis (III)

### The primary endpoint ( Avoid multiple endpoints)

Most RCTs record many variables on each patient. However, at the outset of the trial there are advantages to designing one of these variables as the primary outcome variable (*primary endpoint*).

One of the advantages is that we can focus on the *the primary endpoint when we try to decide how large the trial ought to be*. We aim to gather sufficient evidence to settle questions about the primary endpoint.

This means, the *determination of the primary endpoint in an RCT is quite important*.

42

## Statistical analysis (IV)

### The adequate evidence

Assume that the primary endpoint  $X$  is distributed with mean  $\mu$  in the control group and mean  $\mu + \delta$  in the treated group then the aim is to learn about  $\delta$ , the *effect size*, either by *estimating it* or *testing the null hypothesis* that it is zero.

$$H_0: \delta = 0$$

One could think of "adequate evidence" in terms of making the standard error of some estimator of  $\delta$  small in some sense. However, we face the difficulty of defining "small".

Therefore, adequate evidence is usually thought of in terms of having *adequate power* to test  $H_0$ , say 80% or 90%.

41

## Statistical analysis (V)

### Hypothesis tests

The result of performing a hypothesis test is a probability, universally known as the *P-value* of the test.

$$H_0: \delta = 0$$

Its definition will be

*If the null hypothesis is true, then the P-value is the probability of obtaining a difference between the groups (as measured by the test statistic  $T$ ) as large or larger than that observed.*  $\rightarrow P = \Pr\{T > t \mid H_0\}$

- More informally, if we assume there is no treatment effect,  $\delta = 0$ , then differences between the treatment groups that are very extreme are unlikely to occur by chance and their improbability is reflected in small P-value.

42

## Statistical analysis (VI)

### Hypothesis tests

If a P-value smaller than 0.05 is obtained then this is generally accepted as providing evidence that the null hypothesis is false and we would conclude *there was evidence of a significant treatment effect*.

If a P-value smaller than 0.01 would provide stronger evidence against the null hypothesis.

These conventional thresholds, 0.05 or 0.01, are called *significance levels*.

43

## Statistical analysis (VII)

### Hypothesis tests

If a "non-significant" P-value is obtained, say  $P = 0.235$ , what can be concluded from such a trial?

1) We CANNOT conclude there is no difference between the treatments.

2) We can say the treatment *might have no effect* but it *might have some effect*! Clearly, by itself, this is useless.

Therefore, we MUST design the RCT so that we can avoid making such an unhelpful statement.

- To do this, we must ensure that the size of the trial is sufficiently large to detect clinically important effects, say  $\delta$ .

$\rightarrow$  Calculation of sample size

44

## Statistical analysis (VIII)

### Sample size

To calculate sample size, you must consider the following three parameters:

- Determine the value of  $\delta$ , a *clinically important effect size*,
- Significance level  $\alpha$  ( $= 0.05$ )
- Power  $1 - \beta$  ( $= 80\% \sim 95\%$ )

It should be noted that sample size calculation depends on the test statistic  $T$  to be used for hypothesis testing.

45

## Statistical analysis (IX)

### Summary statistic of primary endpoint

The randomization should have produced groups that are comparable, so the primary comparison is between the primary endpoint  $X$  in the two groups. The summary statistics for the two groups can easily be computed using *mean  $\pm$  SE* (together with *parametric test*) as below:

time	Pravastatin group		Placebo group		t-test p-value
	No.	mean $\pm$ SE	No.	Mean $\pm$ SE	
Baseline	499	4.69 $\pm$ 1.05	510	4.66 $\pm$ 1.09	
Week 26	459	3.47 $\pm$ 0.75	458	4.66 $\pm$ 0.92	
change		-1.22		0.00	< 0.001
% change		-26.01		0.00	< 0.001

46

### Statistical analysis (X)

#### Summary statistic of primary endpoint

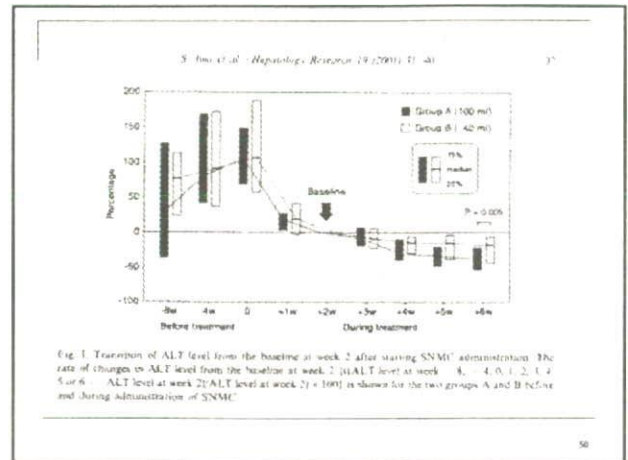
However, if data distribution is highly skewed to the right,  $mean \pm SE$  cannot be used as a summary statistic and instead we should use *median* (25%, 75%) (together with *nonparametric test*) as below:

#### ALT level in a dose finding study of SNMC

time	100ml group		40ml group		Wilcoxon rank p-value
	No	Median (25%,75%)	No	Median (25%,75%)	
Week 2*	45	92.0 (68.5, 119.0)	47	87.0 (71.0, 133.0)	
Week 6	45	50.0 (40.5, 78.0)	46	69.0 (53.0, 85.8)	
% change	44	-38.5 (-54.7, -22.9)	46	-17.3 (-43.8, -7.3)	0.0053

\* Randomization was based upon the ALT level at week 2

49



50

### Goal of Randomized Controlled Trial

#### Reducing Bias

Selection bias	Random allocation
Assessment bias	Masking (blinding)
Analysis of bias	Intent to treat (ITT)
Publication bias	Registry of RCTs

51



臨床研究支援・教育センターが主催した「臨床研究のデザインと進め方に関する研修会（基礎編およびアドバンス編）」報告

分担研究者 尾藤 誠司

研究要旨

国立病院機構が持つ多数の施設と豊富な臨床症例数は多施設臨床研究を行なう上で有利となる基盤を持っている。一方、研究者が自主的な研究を行なう上で、研究計画を立案し、適切な研究計画書を作成することはまだハードルが高い状況にある。本研究事業において、我々研究班は18年度より実地の医療専門職に対し、臨床研究計画を立案する上で必要な疫学的知識、統計学的知識等について、ワークショップ形式でその能力を養成する研修会を行ってきた。さらに、より高い技術を習得するために、「アドバンス研修会」を企画し、適切な研究計画書を作成することができる技術を持った専門職を育成するプログラムを継続した。ワークショップ形式によるプロダクト作成型の研修は、臨床研究技能を身につけるための方法として適切であるといえる。

A. 研究目的

国立病院機構は、145の病院を持ち、20の政策医療分野を中心に様々な種類の疾患を持つ患者の診療にあたっている我が国最大の病院グループである。その意味では、臨床研究、特に多施設共同臨床研究を行なう基盤としては非常に有利な点を持っている。一方、実際に臨床研究を行なう上では必要とされる人的基盤や構造的基盤が必要であり、特に研究者個々の研究計画を立案する能力については我が国においては臨床研究推進における大きなバリアとなっている。臨床研究を計画する上で特に重要な点は、適切なリサーチ・クエスチョンを設定したうえで、リサーチ・クエスチョンに則っ

た最も適切な研究デザインを選択することにあるといわれている。日常診療の中には、あらたな臨床研究の種が数多く存在しているが、実際のところ現場での漠然とした疑問からそれを臨床研究におけるリサーチ・クエスチョンとして明確化し、最適な研究デザインを選択し計画を立てるという作業についての知識や経験が、不足している場合が多いのが現状である。今回われわれは、実際の臨床現場に準じたシナリオを基にリサーチ・クエスチョンを明確化し、最終的には簡易研究計画書を作成できるようになるまでの内容を、講義とワークショップとを交えて行うことを目的とし、研究会を開催した。

## B. 研究方法

平成 18 年、平成 19 年と毎年開催している「臨床研究のデザインと進め方に関する研修会」を、平成 20 年 12 月 11 日から 12 日にかけて、2 日間にわたり開催した。また、本研修会を修了したもの、もしくは、ある程度の基礎的知識・技術を既に習得している医療従事者を対象として、「アドバンス研修会」という形で行うこととし、平成 21 年 2 月 23 日に開催した。いずれも医師・看護師・コメディカルを対象とし、基礎編においてはそれぞれ 22 名・21 名・10 名、アドバンス編においてはそれぞれ 8 名・9 名・3 名と各分野から多くの参加があった。研修会後は、その有用性や満足度について研修会参加者への調査を行ない、研修の反省点等について検討した。

## C. 研究結果

C-1: 「臨床研究のデザインと進め方に関する研修会」

研修会第 1 日目は受付の後、会場となった教室で参加者は 5~6 名程度の小グループに分かれて着席し、まずは互いに自己紹介を兼ねてアイスブレイクから始まった。2 人組でそれぞれ自己紹介をし、今度は他の人々に対して相手の他己紹介を行った。初対面の固い雰囲気をやわらげ、緊張もほぐれたところで講義開始となった。はじめに STEP1『日常の患者ケアにおける疑問をリサーチ・クエスチョンにする』という枠組みで、まずは尾藤による「EBM と臨床研究：臨床の PECO からリサーチ・クエスチョンへ」という講義が行われた。臨床研究の結果として質の高いエビデンスを得ようと試みる場合、まずは質の高いリサーチクエス

ションを打ち立てることが重要であること、日常診療における漠然とした疑問をリサーチクエスチョンへ変換する場合、PECO(P:Patient, E:Exposure, C:Comparison, O:Outcome)の形式を利用すると明確化しやすくなること、また PECO における O:アウトカムは患者の利益になるかどうかの観点で設定すること、などについてのプレゼンテーションであった。ところどころで日常に即した実例を取り入れており、皆聴衆としてだけではなく質問や自身の意見を表明するなどといった形で、積極的に講義に参加していた。続いて実習で、講義内容の実践に移った。「シナリオを用いた PECO とリサーチ・クエスチョンの立案」と題し、実際の現場を想定した臨床シナリオ(今回は、内科病棟勤務看護師における、高齢患者の病棟管理に関する問題を想定)に基づいて、それぞれのグループでリサーチ・クエスチョンを立ててみることを実践した。個々の経験を生かした多彩な提案により、活発な意見交換がなされた。

続いて、STEP2『臨床研究の概念モデルと研究仮説を設定する』という枠組みで、引き続き尾藤により「概念モデルと研究仮説の立て方、及びバイアスについて」という講義が行われた。リサーチクエスチョン(PECO)に基づき、対象母集団(P)、説明因子(E/C)、エンドポイント(O)を設定した概念モデルを作成し、更にそこから研究仮説を立てる具体的な流れが図示とともに説明された。また研究における誤差について、特にバイアスと交絡に関して詳しく述べられた。続く実習では「PECO から概念モデルと研究仮説の設定へ」と題し、各グループで STEP1 の際に作成した PECO に対して

概念モデルを作成し、研究において考慮すべきバイアス及び交絡を挙げるという作業を行った。グループでPECOから概念モデルを作っていくという工程の中で、他職種による異なる見方というものに気付き、非常に刺激的であった、という意見が多かった。

STEP3『臨床研究デザインの種類について』という枠組みでは、財団法人聖ルカ・ライフサイエンス研究所 臨床実践研究推進センター 副センター長の徳田安春氏より、「研究仮説に基づいた臨床研究デザインの決定」と題した大変興味深い講義をいただいた。昨年徳田氏から紹介のあったGATE(Graphic Approach to Teaching EBM)アプローチによる臨床研究デザインの図式化はそのまま踏襲されながらも、GATEの図式が「おでん」に似ていることから、「おでんで臨床研究デザインを考える」というコンセプトでGATEによる研究デザインの整理についての講義が行なわれ、昨年と比較してもより理解が進んだセッションとなった。その後、STEP3『臨床研究のデザインを決定する』をテーマに、「研究仮説に基づいた臨床研究デザインの決定」というワークショップを行ない、研究計画を立案するにあたり、候補となる研究デザインを複数挙げ、その長所・短所および研究事業の実現性について、また何らかの介入を要する研究であればその内容について、検討する作業を各グループで行った。

講習会2日目は、STEP4『取得すべきデータを決定する』では、東京慈恵会医科大学総合診療部・臨床研究開発室 准教授 松島雅人先生から「変数の考え方と種類について」「生物統計の基本的な考え方について」という講義の後、その内容をふまえて「研

究仮説に基づいた変数の決定」の実習が行われた。自分達の作成したPECOと概念モデルに基づいて1次アウトカム・2次アウトカム・説明因子・調節因子を列挙し、アウトカム・説明因子をアウトカム変数・説明変数として定義し、変数の種類を記述し、測定方法を検討するという一連の作業を行った。

昼食を挟み、最後のセッションであるSTEP5『研究計画書を書く』をおこなった。国立国際医療センター研究所 医療情報解析研究部の新保卓郎先生より「研究計画書とケースカードのフォーマット」「臨床研究に関する倫理的勘案事項」について講義をしていただき、実際に「簡易研究計画書の作成」を実習した。この際、患者への不利益、研究事業に関する患者の同意、個人情報保護、の視点から倫理的勘案事項を検討した。最後に2日間の講義および実習を振り返り、自分が何を得たか、今後何をするかについて話し合い、研修会は幕を閉じた。

C-2:「臨床研究のデザインと進め方に関する研修会 アドバンス研修会」

アドバンスコースは前年度から始まった新たな試みで、基礎的な部分に関してはある程度習得されている受講者を対象として、1日コースの研修会という形で開催された。今回は、前年度までのアンケート結果等を踏まえ、実際に研究計画書を作成するための詳細な技術にフォーカスを置いた研修会とした。さらに、リサーチ・クエスチョンを決定する上で必要な文献検索の技術など、基礎編では触れることができなかった部分をカリキュラムに盛り込むこととした。なお、アドバンスコースでは、臨床

シナリオ内で既にリサーチクエスチョン (PECO) が与えられているという形で、それに基づいて実習を行っていくこととした。

午前中は、財団法人聖ルカ・ライフサイエンス研究所 臨床実践研究推進センター副センター長の徳田安春先生を講師とし、東京医療センターシステム研修室にて「文献検索と研究背景を基にしたリサーチ・クエスチョンの吟味」というセッションを行った。同研修室は、14台のインターネット接続が可能な端末をもつ研修室であり、研修生は実際に PubMed などの文献ライブラリに電子的にアクセスし検索実習を行った。

午後のセッションは、「具体的な研究デザインに基づいた研究計画の作成」とし、前編を、東京慈恵会医科大学総合診療部・臨床研究開発室 准教授 松島雅人先生による、「組み入れ基準と除外基準 サンプリングの方法」「変数の詳細な定義」という講義とともに、後半を国立国際医療センター研究所 国際臨床研究センター 医療情報解析研究部長 新保卓郎先生による「スケジュール表と同意説明文書の構造」という講義とともに、実習を行った。実習に際しては、国立病院機構の研究計画書作成のためのひな型を用い、ひな型に沿って簡易計画書や同意説明文書を作成していく、という進行とした。

図表 1 に基礎編におけるアンケート結果を、図表 2 にアドバンス編におけるアンケート結果を示す。基礎編・アドバンス編ともに、参加者からはおおむね高い評価を得ることができた。基礎編においては、「自己負担してでも受講したい」と答えた者は前出席者の 3/4 以上を占めた。講義の内容全

体は、約半数が「価値あり」と返答し、ほぼ全員が「今後の同様の研修は必要か？」という問いに対し「必要」もしくは「必須」と返答した。一方、講義の難易度に関しては、回答者 52 名中 26 名が「やや難しい」と答え、4 名は「難しすぎる」と答えた。

アドバンス編に関しても、約 3/4 は「自己負担してでも受講したい」と返答し、講義内容に関して、ほぼ全員が「価値あり」「きわめて価値あり」と返答した。しかしながら、アドバンス研修においても講義の難易度については 20 名中 11 名が「やや難しい」と、1 名が「難しすぎる」と返答した。

#### D. 考察

本研修会は、国立病院機構内の職員の臨床研究技術の向上を目的とし、平成 18 年度より 3 年度にわたって継続的に行われているプログラムである。リサーチ・クエスチョンの立て方や、臨床研究デザインの適切な選択など、臨床研究を行う上で極めて重要となる骨格部分について、わが国ではまだ医療専門職に十分な教育を提供する機会がなく、その意味においては、研修者は臨床研究の基礎技術部分について知る機会を得るだけでも価値のあることであるであろう。一方で、基礎部分であるとしても、臨床研究の概念的な骨格を理解することは容易ではなく、実際にアンケートの結果からも、内容の難しさを指摘する意見が聞かれている。ワークショップ形式の研修会は、その難しさを場の雰囲気や相互支援、また、事例を基にした作業を通じた上での研修などにより緩和する作用があり、実際にアンケート結果からもワークショップの利点を