

鍼ランダム化比較試験は 実際の試験 (Pragmatic trials), 実際の試験はITT (割付重視) 解析

A randomized controlled trial of acupuncture is a pragmatic trial;
a pragmatic trial is intention-to-treat analysis

SHICHIDO Toshiyuki しちどう としゆき 七堂 利率¹⁾ TAKAHASHI Norihito たかはし のりひと 高橋 則人²⁾

1) 大阪医療学園専門学校 東洋医療教員養成科非常勤講師

2) 明治国際医療大学 加齢鍼灸学教室

はじめに

学会誌を読んでいると論文の質に関して、こうしたらもっと良くなるのに、こうすべきだということがよくある。つまらない間違いで印刷・公表された論文は一生の汚点を残すことになる。一旦公表されると末代まで取り返しがつかない。

これらは事前に誰か識者のアドバイス (校閲の習慣など) があれば改善できることだが、鍼灸の世界では教育機関でも残念なことに、これが機能していない。

では第二の関門である学会誌の査読ではどうかであるが、これも統計的問題については全く機能していないようで、鍼灸関連の学会誌では依然として統計誤用や医学関連知識の誤解が続いている。

そもそもこの2つのゲートが機能していないことがこの原稿を書く動機になっている。

筆者は1992年に『全日本鍼灸学会誌』の統計誤用を調査したが¹⁾、同誌掲載論文の統計誤用は97%に及んだ。その誤用が2011年現在も相変わらず続き、調査結果が活かされていない。90年代前

半、『医道の日本』誌に統計誤用の具体例を著者名公表で指摘したことがあるが、その後一時、ミスの警戒感からか学会誌から有意差検定が消えた。この方法は一定のインパクトがあった。しかし、また忘れやすい国民性か、最近またぞろ頭をもたげてきた。この方法論上の誤りに対する教育・啓発は継続が必要のようである。

万国共通である臨床研究の7つの倫理原則からいうと、科学的妥当性が要求され、「適切な統計手法と正しいデータ」が不可欠である²⁾。いい加減な統計手法や間違ったデータを使うと倫理違反となる。つまり統計誤用は倫理違反となるのである。この点を研究者はもっと自覚すべきである。

過去の文献を引用するというかもしれないで研究する者もいるらしいが、すべての根幹は鍼灸師の教育にEBM (Evidence Based Medicine; 科学的証拠に基づく医療) や統計学が課されていないからである。統計を多用するEBMではあるが、文献の検索は第一優先事項である。筆者からみると、統計を知らないから無関心、あるいは反対するという悪循環に陥っている。その結果、日本の鍼

RCT (Randomized Controlled Trials; ランダム化比較試験) の質が50年間上がっていない³⁾。つまり質が改善されていないのは統計解析の部分と判明している。

そのEBMでは批判的吟味を行い、読むべき論文をチェックして妥当なものの結果を使用する。本論はこの批判的吟味の方法論と言える。

統計誤用対策のためには「誤用に学ぶ」ことが、統計教育の効率化に効果を上げると考えると、本論では最近の論文審査の立場からピットフォール(落とし穴)を実例から指摘し、何故そうなったか、こうすれば良くなるという点を簡潔に説明していき、いささかでも日本の鍼灸のレベルアップに貢献できれば幸いである。

言葉の説明

EBMや統計学等で使われている言葉に対する誤解が多い。まずこれから説明しないと議論が進まない。

1. 鍼RCTはほぼ实际的試験 (Pragmatic trials)

表1のように、試験には实际的試験 (Pragmatic trials: プラグマティック試験) と解明的試験 (Explanatory trials: エクスプラナトリー試験) がある。

实际的というのは日常臨床に近い設定という実際の試験、解明的とは実験室試験に近く、鍼の純粋な効果(例えば施術者の効果を除いた生物学的な効果)を解明するために、シャムでなくプラセボ鍼を対照に置いた試験をいう。

解明的試験の結果は、そのまま日常臨床に当てはめることはできないので、効果 (effectiveness) でなく効力 (efficacy) を表している。いずれ患者サイドでは効果、つまり实际的試験が必要になる。解明的試験は動物実験に近いもので实际的試験への

過程で行われる。

explanatory trialsを説明的試験と和訳している本もあるが、物事は解明してから説明するものではないかと思われ、和訳が誤解を生みやすいのではと思う。

またpragmatic trialsも実践的試験と和訳している本があるが、何を実践するのか、これもわかりにくい。そこで本論では实际的試験と訳している。

表1のとおり、プラセボ・コントロールを使うのが解明的試験、プラセボ (placebo) というのは生理学的には不活性という意味である。鍼は場所を変え、深さを変えてもある程度効果があるのは周知の事実。そこで偽鍼はプラセボ鍼と言わずに、多くはシャム (Sham) 鍼と言っている。シャム鍼なのにプラセボ鍼と称している研究もあり、未だ混乱がある。論文をチェックする立場から言うと、こういう肝腎の言葉を曖昧に使った論文に質の高いものはまずない。

麻酔中の鍼といった特殊な場合、皮内鍼、円皮鍼、提鍼といった毫鍼を除く鍼ではプラセボ鍼が作られる可能性があるが、その他はシャム鍼と言うのが正しい。偽鍼と訳してはシャムかプラセボかわからないので英語を併記することが通常行われている。

そこで鍼RCTではプラセボ・コントロールを使うことはないので、鍼RCTはほとんどが实际的試験 (Pragmatic trials) である。よって鍼RCTの論文にわざわざ英文でPragmatic trialsとタイトルを書くのもナンセンスである。

もう少し2つの試験の違いを表1から説明しよう。

实际的試験では「高い外部妥当性、低い内部妥当性」、解明的試験では「高い内部妥当性、低い外部妥当性」となっている。これは实际的試験では一般RCTと同じく、試験結果が、一般化(自分のところの鍼灸院に適應できるか)しにくいのが、試験は交絡因子の影響が少なく、情報そのものの妥

当性が高いということ。一方、解明的試験ではその逆で一般化しにくく、情報そのものの妥当性が低いということである。

異質患者群を扱い、日常ケアでの施術者スキル (日常臨床に近い) を使うのが実際の試験であり、均質患者群で標準プロトコルでの施術者スキル使用 (実験的要素が強い) が解明的試験である。そこで、より大サンプルが必要なのは実際の試験、より小サンプルで運営できるのが解明的試験となるが、より小サンプルで実際の試験という矛盾した状態が日本の鍼灸RCTの現状である。このサンプルサイズの問題は本シリーズで説明するつもりである。

解析は鍼そのものの効果のみをみる解明的試験では PP (per protocol; アドヒランスの良かったものだけ解析する、プロトコル重視の) 解析、治療者の行為を含む治療効果のみをみる実際の試験では ITT (Intention to treat: 治療を意図した、つまり最初割付けたとおりの) 解析となる。

2. 探索的と検証的試験の区別

鍼RCT論文: 学会誌2001年51 (2), 180頁左側 4 行目に「探索的な試験であるところから per protocol解析を採用し…」とある。

ここで探索的という言葉であるが、探索的 (exploratory) に対立する言葉は検証的 (confirmatory) である。探索的とは「探りを入れる」、検証的とは「証明する」という意味で、データ解析でも前者は探索的データ解析法等を使い、結果を見て仮説をつくることを行う。後者は仮説を確かめるために行う試験で、統計的仮説検定等を使う。

さて、「探索的な試験であるところから per protocol解析を採用し…」という説明は、探索的な試験であるから、脱落者を無視する PP解析を採用したというのは論理的におかしく、意味不明となる。何故なら、ITT解析も PP解析も統計的仮説検定を使うからである。

探索的な試験であるというのはパイロット試験という意味と好意的に解釈しても、小サンプル数

表1 解明的試験 Explanatory trials と実際の試験 Pragmatic trials の区別 (文献⁴⁾ の表1に加筆翻訳)

Explanatory trials	Pragmatic trials
<ul style="list-style-type: none"> • Experimental (実験的) setting • Evaluate efficacy (有効性の評価) • More suitable for acute conditions (急性状態) • Placebo controlled (プラセボ・コントロール) • Patients blinded to minimise bias (患者はバイアスを最小にするためにマスク化) • Aim to equalise non-specific effects (非特異的効果を平均化する目的) • Standardised treatment, simple interventions (標準化治療, 単純介入) • Practitioner skilled for standard protocol (標準プロトコルでの施術者スキル) • Usually short-term follow-up (通常, 短期フォローアップ) • High internal validity, lower external (高い内部妥当性, 低い外部妥当性) • Low relevance/impact on practice (診療のインパクトに低い関連) • Homogenous group of patients (均質患者群) • May manage with smaller sample sizes (より小サンプルで運営) • More commonly used (より普通使用) <p>解析法: PP 解析?</p>	<ul style="list-style-type: none"> • Routine care (日常ケア) setting • Compare effectiveness (効果の比較) • More suitable for chronic conditions (慢性状態) • Not placebo controlled (プラセボ・コントロールでない) • Patients unblinded to maximise synergy (患者は共働するために非マスク化) • Aim to optimise non-specific effects (非特異的効果を最適化する目的) • Routine treatment, complex interventions (日常治療, 複雑介入) • Practitioner skilled in routine care (日常ケアでの施術者スキル) • Often long-term follow-up (しばしば長期フォローアップ) • High external validity, lower internal (高い外部妥当性, 低い内部妥当性) • High relevance/impact on practice (診療のインパクトに高い関連) • Heterogeneous (異質患者群) • May need larger sample sizes (より大サンプル必要) • Less commonly used (普通使用少ない) <p>解析法: ITT 解析</p>

故、p値でなく効果量を求めるというならわかるが検定を使っており、論理的に無理・誤解がある。むしろITT解析で探索的な方法もあり得るし、PP解析で検証的試験もあり得る。試験の性格による区別の知識の混乱・誤解である。統計用語は正確に使ってもらいたい。

3. RCTの結果を患者さんに適応するには 解明的試験の後、実際の試験が必要！

実際の試験 (Pragmatic trials) と解明的試験 (Explanatory trials) は同レベルの対立関係にあると誤解されやすいので、両者の位置関係を述べたい。

鍼がプラセボ効果ではないかという意見の集約は未だない。これは毫鍼のプラセボ鍼が未だできていない証拠でもある。このプラセボ云々は鍼灸師には大切な問題だが、この問題を解決しなくても、実際の試験は可能である。心理効果も含めた鍼効果でも効けばよいわけで、だから鍼RCTでは実際の試験が行われている。

もし毫鍼のプラセボ鍼ができて解明的試験が可能となり、鍼のプラセボが明らかになれば、それは待望の成果となろう。しかし、臨床現場でその結果が使えるかといえ、それはNOである。高い内部妥当性、低い外部妥当性である解明的試験に比べ、高い外部妥当性、低い内部妥当性の実際の試験ゆえ、解明的試験の結果では試験結果が一般化できないのである。最終目標であるRCTの結果を患者さんに適応するというときには、解明的試験を実施してもその後、実際の試験が必要となり、両者は並列でなく前後関係である。

4. 脱落のある場合の解析法はPP解析？ あるいはITT解析？

これまで述べてきたが、解析法には大きく2種ある。

①実際の試験はITT (Intention to treat: 治療を意図した、つまり最初割り付けたとおりの) 解析

②解明的試験ではPP (per protocol; アドヒランスの良かったものだけ解析する、プロトコル重視の) 解析

なお、アドヒランス (adherence) とは支持・信奉・付着の意味で、以前はコンプライアンスと称していたもの。最近はアドヒランスを使うという。鍼RCTはほぼ実際の試験であり、当然解析はITT解析となる。

5. PP解析の問題点

ITT解析するところを、もし間違っ (あるいは意図的に) 脱落者を除いて解析してしまうと、PP解析となる。これだけで発表してしまうと、問題が大きい。その理由は次の2つの視点である。

- 1) 脱落症例の減少：効果のなかった症例をどんどん落としていけば、有効が増えて有意差が出やすいということも起こり、いくらかでも都合のよい結果が得られる。多い脱落例は通常の臨床試験では結果が変わってしまう恐れがあるので、全体の脱落割合の限度は約15%程度までとされている。
- 2) 脱落割合のアンバランス：統計的には群間サンプルの不均衡、つまりどちらかの群で有効あるいは無効例の脱落が多くなれば、ランダム割付が崩れる可能性が出てくる。ランダム割付やランダム抽出はp値計算の大前提であるから、果たして減ったサンプル数で有意差検定をしてもよいのかという問題が出てくる。

6. 論文上の脱落例の扱いはどこに書かれてあり、その他に何をチェックすべきか

脱落例のチェックはRCT論文用チェックリストであるCONSORT⁵⁾を使う。CONSORT関連のフローチャートには脱落の数を記入するようになっている。また、脱落理由まで記載するようになっている (図1)。そしてITT解析したかどうかはCONSORTでチェックするようになっている。

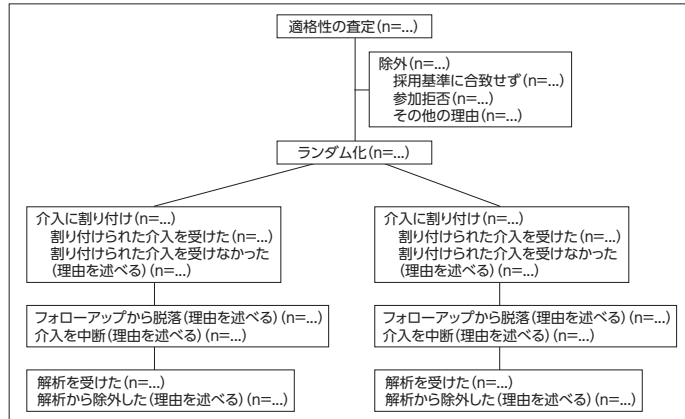


図 1. RTC 論文で必須事項のフローチャート (CONSORT)

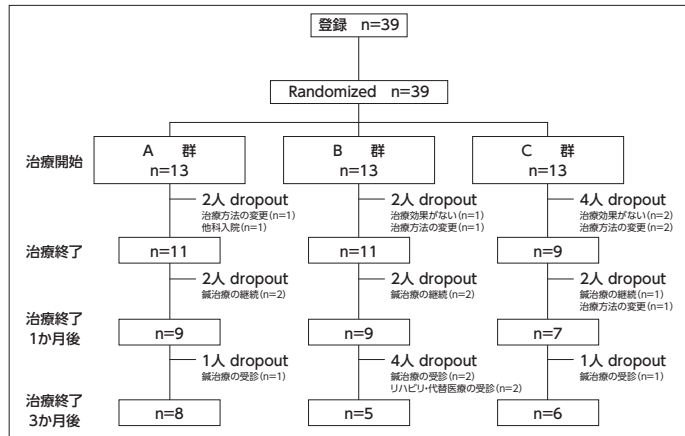


図 2. 『全日本鍼灸学会雑誌』(2009 59 巻 1 号) 掲載の実例 (一部改変)

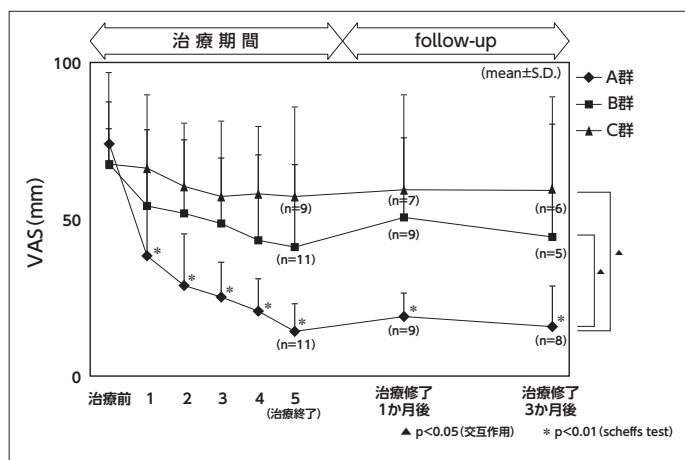


図 3. 腰下肢の痛み (VAS) に対する鍼治療効果

図は各群における VAS の治療効果の平均値を表している。縦軸が VAS の値 (mm) を、横軸が時間経過を示す。グラフは A 群 (◆) と B 群 (■) 及び C 群 (▲) の治療効果の変化を表している。各群の経時的に有意な変化があったもので $p < 0.05$ を ▲ で示し、治療前と比較して有意差があったもので $p < 0.01$ を * で示した。(『全日本鍼灸学会雑誌』(2009) から一部改変)

『全日本鍼灸学会雑誌』も投稿規定から、これに準ずる必要がある。ところがCONSORTのチェック表のITT解析かどうかという項目を、論文査読者が理解しないまま掲載されてしまった実例を図2, 3に示し本論の例題とする。

7. 実際の試験なのにITT解析でなく PP解析という実例

実際の試験なのに、ITT解析していない実例として図2, 3の鍼論文を使って説明していきたい。これは解明的試験と見なしてPP解析らしい。

図3をみると、エラー・バーのSD（標準偏差）が3群でかなり異なる、スタートの初期値ではA：71.2±24.5, B：67.5±19.4, C：70.2±19.2で、A群のばらつき（SD）が一番大きい。それなのにゴールである終了時は、A群だけが最もばらつきが小さくなっている。これは明らかに分布の形状から有効・無効の脱落割合が変化し、除かれた5人に比べ効果のあった8人だけ残ったと推測される。つまり図3からランダム化が崩れている可能性が読み取れる。

交互作用の計算も追跡期間で脱落が終了時より多くなった時点で有意になっているから、群間差は効果なのか脱落のせいかわ不明である。

表2. PP解析とITT解析の差を求めるための表

(最終サンプル数/ 初めのサンプル数)	有 効	無 効 (不変・悪化を 含む)
A群：(8/13)	11 8	2 5
B群：(5/13)	11 5	2 8
C(偽鍼)群：(6/13)	0 6	13 7

注) 量的データ（記載なしのため）を質的データで計算
セル内下段斜体の数字はITT解析で使った例数

図3のデータを表にしたのが表2である。表2の上段の数字は脱落したまま、下段の斜体の数字は脱落例を無効として割付けた時点でのサンプル数。VAS（間隔尺度）の生の値が公表されてい

ないので、「有効」「無効」という質的データで計算した。また当該論文ではVASが改善というが、一体何ミリから改善か、考えないで検定されている。数ミリの改善が臨床的に有効かについては次回に詳しく説明したい。

1) 脱落割合の不均衡の検討：

$$A群：(13-8)/13=5/13=0.3846$$

$$B群：(13-5)/13=8/13=0.6154$$

$$C群：(13-6)/13=7/13=0.4615$$

$$\text{カイ2乗テスト：} p=0.79, C\text{係数}=0.09$$

となり、脱落割合のアンバランスは問題ないように見えるが、サンプル数が少ないために有意差がなくなっただけかもしれない。

2) 再試算

(これらはフリーの計算サイトで計算可能)

① PP解析

表1のデータをVASの生の値が不明なので、質的データで再解析を行ってみよう。帰無仮説は「鍼効果は群間で差がない」となる。

$$\text{結果：Chi Squares}=6.4500 \text{ df}=2 \text{ p}=0.0398$$

(計算フリーサイト：<http://department.obg.cuhk.edu.hk/researchsupport/statstesthome.asp> (香港中文大学産婦人科)で入って、[Frequency tables]をクリックして、[MultiRow/ColTable]を選ぶ。)

なるほど有意差があり、帰無仮説を棄却することになる。

$$\text{市販ソフトMedCalc (1998)でも、} P<0.0005.$$

関連は；C係数=0.627. Cramer's V=0.8045

(関連係数の計算フリーサイト：

<http://vassarstats.net/>から [Clinical Research Calculators]をクリックして「Chi-Square, Cramer's V, and Lambda」に入る。)

鍼の種類と効果は関連有りとなった。

念を入れて正確確率検定は市販ソフト

StatXact-4 (Ver.4.0) で, p値 (Exact) は, $p = 0.0000$ 。

以上, $p < 0.05$ という同じ結果で, 鍼の種類と効果は独立ではない, 鍼群 (図3よりA, B群) が有効ということで, エビデンスの強さは当該論文の結論と同じになる。その上, 論文では論じられていないが, エビデンスの大きさは, 鍼の種類と効果は関連有りとなった。

② ITT解析

次に脱落例で経過不明者を無効例として計算する。

結果は, Chi Squares=1.1166 df=2 $p=0.5722$ 。

3群等しいという帰無仮説を棄却できない。MedCalcでも, $p=0.4876$ 。関連: C係数=0.189, Cramer's V=0.1922

念のため正確確率検定は, StatXact-4 (Ver.4.0) では, p値 (Exact) は, $p=0.6151$ 。

以上, 実際の試験である当該論文では行うべきとされるITT解析では, エビデンスの強さは有意差なし, エビデンスの大きさでは関連無しとなった。

3) PP解析とITT解析の結果が異なるというのは何が起きているのか?

有効・無効の脱落にともなう割合が一定なら, 両者の結果が一致するはずである。簡単に言うと, 無効の脱落割合が大きいと有意差が出やすくなり, 有効の脱落割合が大きいと有意差は出にくくなる。効かない者が落ちていくと有効になりやすくなる。

脱落容認のPP解析でも有効・無効のバランスのよい脱落なら両解析が一致し, アンバランスなら一致しない。ともかく, PP解析の結果だけでは, それがわからない。それでPP解析単独では妥当な解析法とは言えない。

以上, PP解析とITT解析の結果が異なるので偽鍼C群に比べ, A, B群有効とは言えない。もっとはっきり言うと, 両者の解析結果が一致しない場合は, ITT解析を優先し, この場合 (図2, 3の研究) は無効とすべきである。

まとめ

1. 鍼RCTは毫鍼の場合, プラセボ鍼作成は困難なので, 実際の試験 (Pragmatic trials) である。実際の試験は脱落例を無視するPP (per protocol) 解析でなく, 割付重視のITT (Intention to treat) 解析を優先すべきである。
2. PP (アドヒランスの良かった者だけ解析する) 解析を行う時は, ITTとPP解析の両方を行って, 両者の結果が一致し, 両方とも有効となれば結果の妥当性が保証され, それで初めて有効と結論付けるべきで, PP解析だけで有効とすべきではない。

参考文献

1. 七堂利幸: 全日本鍼灸学会誌における統計処理の問題点. 全日本鍼灸学会雑誌, 42(2), 193-8, 1992
2. Ezekiel J. Emanuel, David Wendler, Christine Grady: What Makes Clinical Research Ethical?. JAMA, 283: 2701-2711, 2000
3. 七堂利幸, 井上悦子: 日本の鍼臨床試験 (RTC) の質は上がっているか?. 鍼灸Osaka104, 27(4), 107-113, 2012
4. Hugh MacPherson: Pragmatic clinical trials, Complementary Therapies in Medicine (2004) 12, 136-140
5. 中山健夫・津谷喜一郎編: 臨床研究と疫学研究のための国際ルール集. 60-7, ライフサイエンス出版, 2008

- 1) ☎543-0001 大阪市天王寺区
上本町6-3-31-1111)
- 2) ☎629-0392 京都府南丹市日吉町
明治国際医療大学)