解説

鍼灸臨床効果判定で使う Visual Analogue Scale

七 堂 利 幸*

要旨 計量(metric)データである hard data より、質的(qualitative)データの soft data の重要性は、臨床試験で認識され始めている。その傾向は、quality of life の評価の必要性が最近解かれている事でも分かる。鍼灸は病気でなく病人を治すと言われており、鍼灸研究の評価には、簡便な Visual Analogue Scale(VAS)が、今後多く使われる事が予想できる。そこで、この VAS の作り方、その解析方法、それら実施上の注意点及び、VAS 類似の方法との関連について述べる。

はじめに

鍼灸は 2600 年以前からある治療法である。西洋医学で定義分類されるような疾患を治すのではなく病人そのものを癒すと言われている。それは、患者の訴えに対して処理してきた治療法と言えるのかも知れない。

鍼灸の臨床効果判定には、Feinstein(1977)¹⁾の言うように計量データより質的データ、つまりハードデータよりソフトデータでの評価の重要性が考えられる。血液検査等が正常になっても、主観的な患者の訴えは改善されないことは臨床的に経験されることが多い。主観的な訴えは軽く見られてきたが、現在はこれを重視して評価しようという機運が高まっている。

痛みの評価には、このソフトデータの評価手段の一つである Visual Analogue Scale(VAS)が、よく使われている(Chapman ら 1985; Macrae 1983; Posner 1984; Thomas ら 1985; Max & Laska 1991)^{2~6)}。

最近重要性を帯びてきた quality of life(QOL)

の評価もソフトデータの評価に相当する。現実に癌患者の QOL の評価にも使われている VAS は、Huskisson(1983)らによると簡便で、実用的な方法だと言われている⁷⁾。

McCormack ら(1988)によると⁸⁾、VAS の妥当性の研究は、1968 年から 1983 年にわたり行なわれて高水準の結果がでている。VAS の信頼性については、1973 年から 1983 年までの研究によると、再テスト法や(被)験者間の信頼性試験で確認されている。また、痛みに関する VAS の妥当性の検討も、McCormack ら(1988)によってレビューされている⁸⁾。そして、VAS は既に 1921 年から心理的研究に使われているという。

すでに VAS は、本誌掲載論文でも一部使われているが、その使い方が必ずしも的を得たものではない。今後、VAS は鍼灸の臨床試験その他で多用される可能性がある。そこで、その具体的な作り方、その解析方法、使用上の注意点、類似の方法との関係を紹介する。

1. 長 さ

最もバイアスがかかりにくいのは、Max ら(1991)によると VAS の長さが 10 cm か 15 cm だと言われている⁶⁾。一般的な方法は、10 cm の長さである。間隔はミリメートルだから、100 点単位となる。

作図の時、外端の線の種類と大きさにより、目

* Toshiyuki SHICHIDO 全日本鍼灸学会学術部研究委員会、情報・評価班

Key Words: ソフトデータ、尺度、正規分布、経時測定データ、多重共線性、VAS、RS、GRS、SDS、RAS

の錯覚を生じ、10 cm の長さが見かけ上で変化する可能性がある(田中 1977)⁹⁾。したがって、長軸に対して目盛の長さは、例えば 10 mm 以下と短くし、基線に直角に目盛線を入れる。そして同一試験に使う VAS にはその表示法を統一する。

2. 方 向

VAS の中でもカテゴリーを持った Graphic rating scale(GRS)は、垂直(縦)に置くと、用語の近くにスコアの 75 %が集まり、全長にそって拡がらないで均一な分布をおこさないため、評価が変わると言われている(Scott 1979)¹⁰⁾。また Scott & Huskisson(1976)は、VAS を縦モードから水平モードに変えることによる不履行率(failure rate)は 7 %あるという¹¹⁾。

これは信頼性に関わるので、GRS は必ず水平(横)に置くよう作る。GRS についての説明は 6 節で後述する。図 a , b , c , d 参照。

図 a 横型の VAS

図 b 縦型の VAS

図 c 横型の GRS

自由に [] することができる

全く [] することができない

図 d 個々の患者の特別な評価のための VAS

3. 測定項目と入れるフレーズ

Max ら(1991)によると⁶⁾、カテゴリースケールより、VAS の方が結果の理解が困難であるという意見が多いが、評価者(observer)の注意深い被験者に対する現場の指示が重要になる。

痛みだけでなく、測るべき感覚、反応の程度を的確に表す短いフレーズを尺度の外端に添える。その字体や大きさは同一試験では統一する。

痛みを評価するには VAS で、痛みの程度を聞く方法と、逆に除痛の程度を聞く方法がある。

スケールを紹介するためや、実験の前に適当な質問を用意する。

4. 測定法と評価者および被験者

Scott(1979)によると¹⁰⁾、英国では、専門的職業として臨床評価士という評価技術経験者がいるという。日本では、この臨床評価士という制度がないので、これに該当する仕事をする評価者を置き、必ず実験、試験の内容を知らないマスク(目隠し)された状態で測定に関わる。

患者である被験者自身と研究する側の評価者では、VAS の評価にズレがあるという。特に VAS が 7 cm 以上になると患者の記入する VAS と看護婦の記入する VAS の一致が 7 %と激減するという(水口ほか 1993)¹²⁾。よって自覚症状は患者の訴えであるから、VAS の記入は評価者でなく被験者自身にさせる必要がある。

経時的に測定していく場合、前回の結果を被験者に見せて今回の評価をさせてはいけな

まで現在の時点の評価を独立して行なうべきである。なぜなら、前の回答との間に自己相関が起こる可能性がある(七堂ら 1990)¹³⁾。前の採点を基準に次の採点をした場合、自己相関を起こし、トレンドが発生したりするのである。

繰り返すが、再測定時、以前のデータを被験者に見せたり、それを教えたりして、今の症状を判断させるのは予断を与えることでよくない。

評価者間のバラツキを無くすため、訓練を受けた一人の内容についてマスクされた評価担当者がすべて行なう。多施設の場合は、一人の臨床評価士がカバーすることは容易でない。一つの施設で一つの評価者という体制なら、施設間で差が出ているような場合、後で評価者と被験者の交互作用を調べる必要がある。

VASに付属して口頭でのインタビューの必要があれば、その内容はあらかじめ規格化され、どの被験者にも同じやり方で行なわれなくてはならない。もちろん評価に予見を与える態度や会話は控える。

外来患者を対象とする場合、今日は眼鏡を忘れたから、或は時間がないから自宅へ持ち帰り、次回持参したいという被験者が予想できる。これは帰宅後、家族に記入させる恐れもあり、必ず来院時間内で記入を要請すること。老眼鏡を忘れたという理由だけで、VASの字が見えないという被験者を除外することは、問題であるから、読み上げることは構わないが、読み上げた被験者はその旨メモしておく必要がある。どうしても自宅へ持帰りVASを用いる場合は、初回は当該研究施設でVASを使用し対象患者にマークをつける経験をさせる。その時、VASを用いる方法を十分説明し、初回記入経験を見て自宅へ持ち帰らせる事が可能かどうか見極める必要がある。Scott(1979)によると、これを怠ると50%もの失敗を起すこともあるという¹⁰⁾。

5. 結果の解析

1) 尺度の問題

統計手法は、比率尺度、間隔尺度、順序尺度、分類尺度によって当然異なる(Siegel 1956)¹⁴⁾。しかし、統計手法を選択する前に以下のような問題

がある。

間隔尺度は比率尺度ではないが、大小関係が量的で、1と2の間隔と2と3の間隔を等間隔として問題に出来る。ただ間隔尺度は零点が任意の位置なので、比率は取れない。間違いやすいので例をあげると、間隔尺度には温度表示がある。0点が異なる摂氏0℃と華氏32℃は対応している。両者の2点間の間隔の比は同じである。2点間の距離が知られた大きさということである。

比率尺度は原点として真の0点を持つ、体重でいうと10kgは5kgの2倍である。

本誌には、VASをmmでなくcm単位で評価している例がある。しかし、ここでVASは、長さ10cmの方法を薦めたわけであるから、100mmとして100点単位測定とする。厳密には個々の試験毎に何ポイントがよいか、予備試験でスコアの分布を調べる必要がある。被験者間デザインと被験者内デザインではそれらが異なる可能性があるので、試験デザインによる予備試験が必要である。

測定値を間隔尺度とみなし、t検定などをかけるということが日本では当たり前のように行われている。しかし、本来これは「強い、中等度、弱い」といった順序尺度である。順序尺度は数値の大小関係を表しているが、間隔が意味を持っていない。順序尺度は加減乗除ができないのである。つまり鍼灸効果で今後頻用されると考えられる順序尺度というのは、マラソンの順位と考えると理解しやすい。1位の人は2位の人の2倍速かったわけでないということである。これらは尺度の等比性と等距離性の問題である。

従って平均値や標準偏差を求めることは、厳密には問題がある。それではVASを用いた計算は誤用ということになる。しかし、これは便宜的に順序尺度を間隔尺度とみなしているということである。例えば、鍼の研究にも使われているが、情意的意味体系を探ると言われているOsgood(1962)のSD(Semantic differential)法でも順序尺度を間隔尺度と見なして計算している¹⁵⁾。これは経験的に近似できるということで大目にみているわけである。見なしているというのは、誤解されやすいので説明すると、順序尺度を間隔尺度と見なした場合、間隔尺度までの検定法が使えることになり、統計

量の種類は多くなる。解析する方にとっては便利になる。

しかし、実用上問題無いと言っても、厳密には結果的に誤りを冒すかも知れない。これは実用主義と理論主義の対立である。両者の折衷案として、正規化順位法というのがある。これは順位を正規スコアに変換し、各サンプル毎に推定して評定し、解析する方法である。計算方法は佐藤(1978)を参照願いたい¹⁶⁾。ただし、この方法は計算が面倒になり実際は一般的でないという。

以上の理由で、VAS から得た数値は間隔尺度と見なすことが可能である。

VAS から得た数値を処理して疑問点が生じるようなことがあれば、単に機械的に計算するのではなく上述の点を吟味する必要がある。

2) 分布型の問題

次に、分布についても正規分布するという保証は無い。予備試験でスコアの分布を調べる必要性はここにもある。分布の適合度はその検定や、確率プロット法である P-Q プロットなどを使えば良い(脇本・垂水・田中編 1987)¹⁷⁾。正規分布が証明されない場合、例えば Macrae(1983)のように³⁾、一様分布には arcsine 変換を薦めている人もいる。これは、0 から 10 単位までの変化と 40 から 50 単位までの変化が同一強度であるという保証が無いものを調整し、変換しようというものである。

検定は尺度の種類によって変わる。間隔尺度と見なして独立 2 資料の比較で t 検定を使うときは、母集団が正規分布で等分散という前提を思い出す必要がある。まず初歩的には F 検定で分散の検定をして、等分散を確認してから t 検定を使うべきである。母集団が正規分布で分散が等しくなければ、Cochran-Cox 法や Aspin-Welch 法を使う。正規分布でないときは、測定値の変換の他にはノンパラメトリック(分布型によらない)検定法を考える。特にデータが間隔尺度の場合、Siegel(1956)は既に対応のある 2 群及び独立 2 群の為の randomization test を薦めている¹⁴⁾。これを使うと母集団や標本集団の概念は必要ない。randomization test は非常に現実的な方法である。言い換えると、この方法を使う時は、無作為抽出という前提条件は必

要なく、無作為割付という前提条件だけでよいのである。

3) 検定法を使う時の諸問題

VAS の検定で薦めたいのは、厳密性からいうと順序尺度に合った検定法、例えば累積 χ^2 法を使ったり、あるいは Wilcoxon の検定などのノンパラメトリックな方法の使用である。また、間隔尺度にあった検定法も実用上問題はないので 2) で述べたように randomization test を薦めたい。

経時測定データの場合における各時点の群間比較には、Student-Newman-Keuls のような多重比較法を使う。

薬物の高血圧に対する大規模な多施設無作為化比較臨床試験における二群以上の比較の問題で、VAS を用いた多変量の複雑なデータを簡単に処理する方法を提示した Solomon(1989)の文献もある¹⁸⁾。これは、Fisher の Z 変換(等相関の検定)を利用したもので、医学データの処理法として注目してよい。

著効、有効、無効、悪化という 4 段階で最初にデータを取ったなら、解析もそのまま 4 段階で行うべきである。それを解析時に 2 段階や 3 段階に併せて計算するのは誤りで、尺度合わせ(scale merging)といって有意差がでやすくなるので注意が要る。

鍼灸の研究では、同一被験者内での繰返し測定が頻用される。McCormack ら(1988)によると⁸⁾、VAS はまた、被験者間の測定より、この繰返し測定に適しているという。これは繰返し測定の利点が活かせるということである。つまり繰返し測定デザインでは個体内変動が少ないということで背景因子の均一化が活かせるということである。この解析方法は、繰返し測定の分散分析や混合効果モデルなどが使用できる。混合モデルとは母数因子と変量因子が同時に扱われるモデルをいう。ただ母数モデル、変量モデル、混合モデルのデータの形は全く同じで、異なっているのはデータの見方、検定方法、そして結果の解釈の仕方である。つまり結果的に限界水準が影響を受ける。

同一症例から VAS で、多くの項目を同時に評価したい時は、個体の独立性を考える必要がある。

一つめは、同じ対象から繰返し測った測定値をそれぞれ独立とみなすプールする誤り (pooling fallacy) を犯さないよう注意がいる。鍼研究で本誌にもこの誤りが散見される。検定時、多項目検定といった多重性の問題が生じる。この場合は、タイプ1のエラーを考慮した方法を選んで検定することが重要である。多重比較を使おうとする場合は、事前のサンプルサイズの設定も忘れないようにしたい。サンプルサイズに関する文献と市販ソフトウェアのレビューは七堂(1995)を参照のこと¹⁹⁾。

第二に、同一症例から多項目をとる評価では項目毎の独立性が損なわれ、例えば「馬から落ちて落馬して」というような説明変数間に強い相関関係が生じ、推定値に大きな標準誤差を持つ多重共線性 (multicollinearity) の問題を犯さないようあらかじめ独立した項目を選ぶことが必要である。多重共線性のあるモデルでの対処法は、Sall(1981)が分かりやすい²⁰⁾。

Keats(1971)によると²¹⁾、ある人のVASによる評価と他の人の評価とは、比較出来ない。このことは認識しておくべきである。

VASの利点は測定値が連続的と見なして計算するから、カテゴリースケールの時に起るかも知れない重大な境界エラー、言い換えると分類することにより消失する情報からくるエラーは避けられるかもしれない。

6. 類似の方法

単純記述スケール(SDS)は、文字だけでその症状の程度を順序尺度で表すもので、鍼の研究でも作られたことがあった。それを図形化したのが、図eのSDSである。ただし、Scott & Huskisson(1976)によると、これはVASより鋭敏性(sensitivity)で劣る¹¹⁾。

除痛効果を見た、VASとカテゴリースケールはよく相関し、痛みの程度を見たVASとカテゴリースケールで測ったものともよく相関している。ただ、解放スケールでは、VASの方がカテゴリースケールよりもより鋭敏であるという(Maxら1991)⁶⁾。

VASは、RS(Rating scales)の一種であるという見方が一般的であるが、Brantley & Bruce

症状：

著効
有効
無効
悪化

図 e 単純記述スコア(SDS)

(1986)は両者を区別して並列に分類記述している²²⁾。

Scottら(1976)¹¹⁾やMacら(1984)²³⁾のGRSは、VASに記述的用語を載せたものである。GRSは、VASの一種と考えられる。GRSの実例は図cを参照のこと。

GRSでは、使用時スコアの分布を均一にするため縦型より、横型がよいと言われていることは既に述べた。

Macら(1984)によると²³⁾、VASをエンドポイント²⁴⁾(endpoint: 臨床試験で使われる用語。試験対象で観察された時、その対象では試験中止か治療変更となる事象。)以外の他の評価項目に利用するには、患者にとって必ずしも容易でない。その点、GRSは、カテゴリーを持ったVASであるため、使いやすいという。Scottら(1976)によると¹¹⁾、VASとGRSは、単純な記述的痛みとの相関は高い。これらは伝統的な単純記述的痛みスケールより、もっと鋭敏である。

VASは、Rating Analogue Scale(RAS)との相関もよい。図aや図cなども全てRASと言えなくもないが、区別するとRASとは、例えば時計盤表示のスケールになっていると言ったものである(図f)²⁵⁾。

Brantley & Bruce(1986)によると²²⁾と、最も広く使われているRSは、Verbal Rating Scale(VRS)である。数表現あるいはVRS単独では、患者の痛み強度を十分評価出来ないという。またVRSは、尺度カテゴリー間の等間隔の推定が出来

図 f 痛み評価で使われている RAS〔折笠 1995〕

ないが、VAS ではこれらが可能であると言われている。Ohnhaus ら(1975)によると、VRS と VAS も高い相関を示している²⁶⁾。

VAS は、5 評価点の VRS より、鋭敏だと言われている。また、VAS は他の痛みの評価より鋭敏性が高いと言われている。

10 項目分類の VAS は、Behaviour Rating Scale(BRS)との相関性も高い(Meister 1985)²⁷⁾。BRS と類似の Behaviour Anchored Rating Scale(BARS)の実例や Behaviour Observation Scales(BOS)の内容とそれらの比較については、Meister(1985)参照のこと²⁷⁾。

McCormack ら(1988)によると⁸⁾、Visual Analogue Mood Scale(VAMS)とは、特殊な質問を使うもので、例えば「今のあなたの気持ちは？」あるいは「あなたの最も良い気分は？」と言った具合である。

VAS によるうつ病(Deprssion)とうつスケールとの相関、VAS による不安スケールとの相関は McCormack ら(1988)にある⁸⁾。精神神経薬理学(psychopharmacology)で使われるその他多くの種類の RS について詳しいのは、Vanriezen & Segal(1988)の文献である²⁸⁾。

VAS で捉えにくい情動的、感覚的痛みの性質には、水口公信ほか(1993)のいう「今の痛みを一言でどのように表現しますか」と尋ねる心理テストの文章完成法を VAS と併用することも考慮してよい¹²⁾。

謝辞

本論をまとめる機会を与えてくださり、ドラフトの段階で多くの有益なコメントを頂いた明治鍼灸大学 川喜田健司先生に感謝します。

参考文献

- 1) Feinstein A. R.: Clinical biostatistics XLI. hard data, soft data, and the challenges of choosing clinical variables in research, *Clinical Pharmacology and Therapeutics*, 22(4), 486-498, 1977.
- 2) Chapman C. R. et al.: Pain measurement; an overview, *Pain*, 22, 1-31, 1985.
- 3) Macrae K. D.: The interpretation of pain measurements, In Ronald Melzack(Eds.) *Pain measurement and assessment*, Raven Press, pp21-25, 1983.
- 4) Posner J.: A modified submaximal effort tourniquet test for evaluation of analgesics in healthy volunteers, *Pain*, 19, 143-151, 1984.
- 5) Thomas P. Guck et al.: Multidisciplinary pain center follow-up study; evaluation with a no-treatment control group, *Pain*, 21, 295-306, 1985.
- 6) Max M. B., Laska E. M.: Single-dose analgesic comparisons, In M. Max et al.(Eds.), *Advanced in pain research and therapy*, Raven Press, Vol.18, pp55-95, 1991.
- 7) Huskisson E. C.: Visual analogue scales, *Pain measurement and assessment* Edited by Ronald Melzack, Raven Press, 33-37, 1983.
- 8) McCormack, E., et al.: Clinical applications of visual analogue scales: a critical review, *Psychol. Med.*, 18, 1007-1019, 1988.
- 9) 田中良久: 心理学測定法, 東大出版, (1977), p. 27~29.
- 10) Scott J.: 誰が評価測定するのか, 抗炎症剤の薬判定, 方法と倫理, 医学書院, (1979), p.87~91.
- 11) Scott J. and Huskisson E. C.: Graphic representation of pain, *Pain*, 2, 175-184, 1976.
- 12) 水口公信ほか: がんの痛み評価, *臨床麻酔* 17 (6); 721~728. (1993)
- 13) 七堂利幸ほか: 中国武術の型(表演)試合における審判方法の問題点, *武道学研究* 22(3); 57~64. (1990)

- 14) Siegel S.: Nonparametric statistics for the behavioral sciences, McGraw-Hill, 1956.
- 15) Osgood C. E.: Studies on the generality of affective meaning system, Amer. Psychologist., 17, 10-28, 1962. (吉田正昭訳, 計量心理学リーディングス, 誠信書房, 1968).
- 16) 佐藤 信: 官能検査入門, (1978), p.70~76.
- 17) 脇本・垂水・田中編: パソコン統計解析ハンドブック I 基礎統計編, 共立出版, (1987), p.204~225.
- 18) Solomon P. J.: Analysis of patient self-assessment visual analogue scales, Applied Stochastic and Data Analysis, 5, 153-164, 1989.
- 19) 七堂利幸: サンプルサイズの求め方に関する文献・ソフトのレビュー, 全日本鍼灸学会誌投稿中
- 20) Sall J. P.: SAS Regression Applications, 1981. (新村秀一訳, SAS による回帰分析の実践, 朝倉書店, (1990), p.75~88.)
- 21) Keats J. A.: An introduction to quantitative psychology, Wiley & Sons, 1971. (島津, 石井訳, 計量心理学, 誠信書房, 1976)
- 22) Brantley P. J., Bruce B. K.: Assessment in behavioral medicine, Ciminero A. R., et al. Ed.; Handbook of behavioral assessment, Wiley-Interscience, 673-721, 1986.
- 23) Mac W. Helt and Scott R. Parker: An experimental basis for revising the graphic rating scale for pain, Pain, 19, 153-161, 1984.
- 24) 佐久間 昭: 臨床試験での統計用語, 国際化時代の新薬薬効評価, 新医薬品研究開発フォーラム 5 ミクス; 94~117. (1993)
- 25) 折笠秀樹: 臨床研究デザイン, 真興交易医書出版部, (1995), p.104.
- 26) Ohnhaus E. E. and Adler R.: Methodological problems in the measurement of pain; a comparison between the verbal rating scale and the visual analogue scale, Pain, 1, 379-384, 1975.
- 27) Meister D.: Behavioral analysis & measurement methods, Wiley, 320-332, 1985.
- 28) Vanriezen H. and Segal M.: Comparative evaluation of rating scales for clinical psychopharmacology, Elsevier, 1988.
(〒543 大阪市天王寺区上本町 6-3-31-1111)

The Visual Analogue Scale used in Judging Clinical Effectiveness of Acupuncture and Moxibustion

Toshiyuki Shichido

Information and Evaluation Group (IEG)

Research Committee, Scientific Department

Japan Society of Acupuncture (JSA)

Summary

The relative importance of soft, or qualitative, data over hard, or metric, data is now beginning to be recognized in clinical trials. This tendency can be seen in the recent importance attached to evaluating quality of life. It is said that acupuncture and moxibustion are used to cure the patient, not the illness. This makes it likely that the simple and convenient Visual Analogue Scale (VAS) will come into wide use henceforth in evaluating research in acupuncture and moxibustion. Therefore, we would like to take this opportunity to discuss how to make a VAS, how to analyze it, points to look for in implementing it, and the relation between VAS and similar methods.