Bulletin of Toyohashi Sozo University
2013, No. 17, 1–8

筋力テストの信頼性評価を使用したShroutとEliasziwの信頼性係数の比較

柴田賢一[1) 2)]　森嶋直人[1)]
宮原英夫[2)]

抄　録

筋力テストの信頼性の評価法としてShroutらの日内信頼性係数（ICC(1,1)）と日間信頼性係数（ICC(2,1)）による評価が普及している．しかし，この方法は理学療法でよくみられる1日に複数回の測定を複数日行う形式の筋力テストの信頼性評価にそのまま適用できない．Eliasziwらは，このような形式のデータをそのまま使用し，日内と日間の信頼性係数を同時に与える別の信頼性係数を提唱している．我々は自身で行った足関節底屈筋の筋力テストで得られたデータを使って，Shroutらの信頼性係数とEliasziwらの信頼性係数を算出し，筋力テストの信頼性評価における両者の特徴を比較した．テストの日内信頼性では，検査日別に評価するICC(1,1)が，各検査日の個別の特徴を示すことができたが，Eliasziwらの日内信頼性係数は調査全期間のデータを一括して取り扱い，検査日別の評価は与えなかった．テストの日間の信頼性の評価では，ICC(2,1)がEliasziwらの信頼性係数よりもやや高い値を示したが，ほぼ同等の値を与えた．一方，他の筋力テストと比較する際には，繰り返し数の情報がないと公平な比較ができないので，Eliasziwらの日間信頼性係数が優れていると考えられた．このような特徴が認められたが，今回対象とした筋力テストの信頼性評価では，どちらの信頼性係数で評価しても，日内信頼性，日間信頼性共によく似た値が得られ両者の判定に大きな差は認められなかった．

キーワード：テストの信頼性，級内相関係数，信頼性係数，分散分析

1．はじめに

理学療法で筋力トレーニングを長期間続けて筋力の増強を試みたとき，その効果の判定は，トレーニング開始時と終了時に適当な筋力テストを行ってその結果を比較すればよいと考えられるが，筋力の増加が小さい場合には，測定の感度だけでなく使用するテストの信頼性（再現性）が十分に高い必要がある．信頼性が高くないと，仮に判定時の筋力が開始時の筋力よりも増加していたとしても，それがトレーニングの効果によるものなのか，たまたま大きい値が得られたのか判断できない．

1）豊橋市民病院　リハビリテーションセンター
（Department of Rehabilitation Medicine, Toyohashi Municipal Hospital）
2）豊橋創造大学大学院　健康科学研究科
（Graduate School of Health Sciences, Toyohashi Sozo University）

テストの信頼性は，測定の対象がたとえば血清中のNa濃度のように検体として取り出せるものであれば，標準となる試料を用意し，何回か測定を繰り返して得られた一連の測定値のばらつきの大小によって評価することができる．しかし，筋力テストのように人間を対象として測定が行われる場合には，得られる測定値は機械的な測定誤差だけでなく，測定者によっても変化すると考えられるので，単純に測定誤差をみるだけでテストの信頼性を評価することは出来ない．

このような場合，古典的テスト理論[1)]では，得られた測定値が真の筋力と測定誤差の和だけで成り立っていると考えずに，真の筋力とそれ以外の要因の和で成り立っているというモデルを考えて，仮定した真の筋力の分散と測定値の分散の比で信頼性係数を定義し，それを使って信頼性を評価している．特にShroutら[2)]が，級内相関係数（Intraclass Correlation Coefficient; ICC）や分散分析表に現れる統計量と対応させてICC(1,1)，ICC(1,k)，ICC(2,1)，ICC(2,k)，ICC(3,1)，ICC(3,k)と6種類に整理した信頼性係数は，理学療法の分野で広く利用されている[3, 4, 5)]．

最も基本的な信頼性係数は，一人の検者が複数の被検者にそれぞれ複数回の測定を実施して得られた測定値を，モデルに当てはめて算出した検者内信頼性係数である．Shroutらの分類のICC(1,1)に相当し，その算出には，1元配置分散分析が利用される．このモデルを拡張して，一人の検者が各被験者当たりm回行っていた測定を，m人の検者が1度ずつ分担して行った場合に置きかえてモデルを構成し，ICC(1,1)で一まとめにしていた筋力と検者の影響を分離して調べることもできる．このモデルに基づく信頼性係数は検者間信頼性係数と呼ばれ，繰り返しのない2元配置分散分析が利用される．これはShroutらの分類のICC(2,1)あるいはICC(3,1)に相当する(注1)．

Shroutらは，テストの信頼性に影響する要因として検者を取り上げて説明しているが，モデル上では，検者でなくても，検査の場所であってもよいし，検査日であってもよい．この小論では以後，検者を検査日と置き換えて検討を進めることとする．

理学療法の臨床で行われる筋力テストでは，同じ日のテストでも，1度だけ筋力を測定するのではなく，複数回繰り返して筋力を測定し，それをまとめて1セッションのテストとすることが少なくない．ICC(2,1)では各検者が1被検者当たり1度だけしか測定を行わないことを前提にしてモデルが構成されているので，1被検者当たり複数回の測定を行ったときには，複数個の測定値を平均値や中央値など一つの代表値に置き換えてからでないとICC(2,1)は算出できない．また，複数個の測定値でテストが表されている場合，ICC(2,1)を使用すると，原データの持つ情報の一部が失われる．

Eliasziwら[6)]は，各被検者につき1日1個だけでなく，測定した数だけ測定値を利用できるように拡張したモデルと，それに基づく信頼性係数を提案している．1日のテストで反復測定を行う場合には，Eliasziwらの信頼性係数の適用も考えられるが，理学療法の分野では，ほとんど使われていない．我々は，自身で行った足関節底屈筋の筋力テストの信頼性をShroutとEliasziwの2種類の信頼性係数を使って評価し，それぞれの信頼性係数の利点と欠点を比較検討した．

2. 使用データと解析方法

23歳から35歳までの健常男性10名の膝伸展位における右足関節底屈筋の等尺性収縮のトルクをBiodex System 4（Biodex社製）を用いて測定した．5秒間の休憩をはさんで3秒継続する最大収縮を計5回繰り返させ，1セッションのテストとした(注2)．同様のテストを1日に1セッション，3日間計3セッション実施した．すべてのテストは同じ検者が担当し，日を改めて測定する際には，被検者の座る位置や椅子の設定に日差が生じないように注意した．テストの詳細については別に報告した[7]．

得られた一人当たり延べ15個の測定データを検査日別に整理して**表1**に示した．各検査日別の表の右から2番目の列と最右列にそれぞれ被検者別の平均値と標準偏差を，最下行に検査順ごとの平均値を示した．また，1セッション内の繰り返し数をm，被検者数をn，検査日数をtとして全データの構造を**表2**に示した．今回のデータでは，m = 5, n = 10, t = 3である．

筋力テストの信頼性係数の算出は次のように行った．先ず各検査日別に分割したデータを対象に，定義(注3)に従い1元配置分散分析表を使って日内信頼性係数：ICC(1,1)を算出した．日間信頼性係数：ICC(2,1)は，各被検者のi日目の測定値として，i日に行われた5回の測定の平均値を使用し，定義[4]に従い「繰り返しのない2元配置分散分析表」を使って算出した．

Eliasziwらの日内信頼性係数（σ_{intra}）と日間信頼性係数（σ_{inter}）は，定義(注4)に従い，「繰り返しのある2元配置分散分析表」を利用して算出した．その際，**表2**の構造に整理した3日分の全データを1度に使用した．

信頼性の定性的な評価は得られた信頼性係数の値をLandisの基準[8] (注5)に当てはめて行った．ICC(2,1)の95%信頼限界は，SPSS（statistic 19）を利用して算出した．筋力の測定は豊橋創造大学倫理委員会の承認（研究課題番号21102011）を受けた後，被検者に実験の目的，研究内容・方法について十分に説明し，口頭による同意を受けた上で豊橋市民病院リハビリテーションセンターで実施した．

3. 結果

3. 1. Shroutらの日内信頼性係数（ICC(1,1)）による評価

ICC(1,1)は，第1日目，2日目，3日目の各テストに対してそれぞれ0.86，0.89，0.86という高い値が得られた．これらは，いずれもLandisの基準の最高レベルに該当した．この結果，ICC(1,1)による評価では，同一日内に行われた筋力テストの信頼性は，すべての日で高いと判定された．

3. 2. Shroutらの日間信頼性係数（ICC(2,1)）による評価

ICC(2,1)は0.55（95%信頼区間；0.18 ～ 0.84）であり，点推定値は先に算出した各検査日ごとのICC(1,1)のいずれよりも低下した．0.55という値はLandis基準の上から3番目の

レベルに該当し，筋力テストの日間信頼性は，まだ十分でないと判定された．

95%信頼区間は広く，Landis基準の5段階のレベル，低いから良好まですべてのレベルに亘っていた．

3. 3. Eliasziwらの日内信頼性係数（σ_{intra}）による評価.

σ_{intra}は0.87で，Landis基準では最高レベルに該当し，同一日内に行われた筋力テストの信頼性は高いと判定された．各検査日別に求めた3個のICC(1,1)と比較すると，検査日別でも，平均でもよく一致していた．

3. 4. Eliasziwらの日間信頼性係数（σ_{inter}）による評価.

σ_{inter}は0.49でLandis基準では，上から3番目のレベルに該当し，筋力テストの日間信頼性は，まだ十分でないと判定された．0.55（0.18 ～ 0.84, 95%CI）を得たICC(2,1)と比較すると，約1割低下したが，ICC(2,1)の95%信頼限界に含まれていた．

4. 考察

4. 1. 日内信頼性係数

ICC(1,1)は各検査日個別に日内信頼性係数を求めるので，テストの信頼性が検査日別に評価できる．したがって，測定手順や担当者が代わるなど，各被検者の測定値のばらつきが検査日によって大きく変わることが予想されるような場合には，有用である，

一方，Eliasziwらの信頼性係数は原データの構造と「繰り返しのある2元配置分散分析」との対応が理解し易いという特徴がある．一旦分散分析表が用意できると，日内信頼性係数σ_{intra}と，日間信頼性係数σ_{inter}を同時に算出できる．ICC(1,1)が，各検査日ごとに日内信頼性係数を算出するのに対し，σ_{intra}は全検査日のデータを一括して日内信頼性係数を算出するので，全検査日に対する共通の日内信頼性係数が1個の値に集約されるという利点がある．特にσ_{intra}が高い場合には，各検査日のICC(1,1)がすべて高いことが1個の指標で表示される．今回，評価の対象とした筋力テストでもσ_{intra}が0.86と高く，全てのICC(1,1)も高いことが予想でき，実験結果も予想と一致した．しかし，σ_{intra}が低い場合には，改めて各検査日で個別にICC(1,1)を求めて低下の原因を検討する必要がある．また，各被検者の測定値のばらつきが検査日によって変化しないというモデルの前提が成立していない可能性もあり，検討が必要になる．

4. 2. 日間信頼性係数

今回，我々の自験例を対象として算出した日間信頼性係数はICC(2,1)が0.55で，0.49であったσ_{inter}よりも1割程度高値を示した．このように平均値を使用する前者の信頼性係数が後者のそれよりも高い値を示す傾向が知られているので[6]，報告に当たっては，どちらの信頼性係数を使用したか付記することが望ましい．理学療法の分野では，1セッション内の

複数の測定結果を平均値で代表させることが少なくないので，日間信頼性係数をICC(2,1)で評価することは実用的である．1セッション内の個々の測定値の情報を使ったσ_{inter}で評価される筋力テストは，平均値で評価される筋力テストよりも概念があいまいであるが，今回の比較では大きな差が認められなかった．ICC(2,k)は，1セッション内の複数回の測定に対応する信頼性係数ではないので，利用できない．

一方，異なった筋力テストの日間信頼性係数を比較する場合には，ICC(2,1)のように筋力テストの代表値だけでなく，代表値が何個の測定値から得られているかという情報が必要になるので[6]，σ_{inter}を使用して比較することが望ましい．

5．結語

我々は足関節底屈筋の筋力テストで得られたデータを使って，Shroutらが提示した信頼性係数（ICC(1,1)とICC(2,1)）ならびにEliasziwらが提示した信頼性係数（σ_{intra}とσ_{inter}）を算出し，筋力テストの信頼性評価における特徴を比較した．テストの日内信頼性の評価では，検査日別に評価するICC(1,1)が，各検査日別に信頼性の高さを示すことができたのに対し，σ_{intra}は，調査全期間における日内信頼性の評価を行うことはできたが，各検査日別にそれぞれの特徴を示すことができなかった．テストの日間の信頼性の評価では，ICC(2,1)とσ_{inter}がほぼ同様の評価を与えた．異なった筋力テストの日間信頼性係数を比較する際には，σ_{inter}の使用が望ましいと考えられた．

注1．ICC(2,1)とICC(3,1)との違いは，分散分析の理論で，検者を特定の人々であるとみなすのか，それとも多数の候補者の中から，たまたま選ばれた人々であるとみなすのかに由来するが，この論文ではICC(2,1)だけを取り上げて話を進めることにする．

注2．この論文では1回のテストという表現を，1度だけの測定に与えるだけでなく，同じ目的のために何度か繰り返して行われた一連の測定（1セッションの測定）に対しても使用した．

注3．ある特定の検査日iにおける日内信頼性係数（$\rho_{intra}(i)$）は，次の式を使って推定する．

$$\rho_{intra}(i) = (MSS(i) - MSE(i))/(MSS(i) + (m-1)\times MSE(i))$$

MSS(i)とMSE(i)は，検査日iの測定値を使って行った1元配置分散分析表（**表4**）中の被検者間平均平方と被検者内平均平方であり，mは測定の反復数（=5）である．このモデルはShroutらの報告のCase 1に相当し，$\rho_{intra}(i)$は，検査日iにおけるShroutらの検者内信頼性係数ICC(1,1)に対応する．

注4．Eliasziwらによるテストの信頼性の評価指標は，繰り返しのある2元配置分散分析表（**表3**）の期待値の欄に示される被検者間分散（σ_A^2），測定日間分散（σ_B^2），交互作用分散（σ_{AB}^2），誤差分散（σ^2）を使用して，次のように定義される．

同一日内の信頼性係数（$\rho_{日内}$）：

$$\rho_{日内} = (\sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2)/(\sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma^2)$$

日間の信頼性係数（$\rho_{日間}$）：

$$\rho_{日間} = \sigma_A^2/(\sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma^2)$$

$\rho_{日内}$の推定値（ρ_{intra}）は，分散分析表中の被検者間平均平方（MSS），日内平均平方（MSE），日間平均平方（MSR），交互作用項の平均平方（MSSR）と，テストの反復数（m=5），被検者数（n=10），日数（t=3）を次の式に代入して算出する．

$$\rho_{intra} = A/(A + MSE)$$；

$$A = (MSS - MSSR)/(m\times t) + (MSR - MSSR)/(m\times n) + (MSSR - MSE)/m$$

上式の右辺の3項は，それぞれσ^2_A，σ^2_B，σ^2_{AB}，の推定値である．また，MSEはσ^2の推定値である．

同様に$\rho_{日間}$の推定値（ρ_{inter}）は，2元配置分散分析表中の統計量とそれから求められたAを次の式に代入して算出する．

$$\rho_{inter} = (MSS - MSSR) / (m \times t \times (A + MSE)).$$

注5．Landisらの5段階判定基準（Landis基準）

信頼性係数の高低に応じてテストの信頼性を，低い（slight　0.0–0.20），不十分（fair　0.21–0.40），まあまあ（moderate　0.41–0.60），かなり良好（substantial　0.61–0.80），良好（almost perfect　0.81–1.00）と大別する基準である．

参考文献

1．南風原朝和：古典的テスト理論，テストの信頼性と妥当性．松原望（編）統計学100のキーワード．弘文堂，2005，pp.122–127.

2．Shrout PE, Fleiss JL: Intraclass correlations: uses in assessing rater reliability. Psychological Bulletin. 1979;86(2):420–428.

3．Bartko, JJ: The intraclass correlation coefficient as a measure of reliability. Psychological Report. 1966;19:3–11.

4．対馬栄輝：検者間・検者内信頼性係数．SPSSで学ぶ医療系データ解析．東京図書，2005，pp.195–214.

5．Matthews, DE, Farewell, VT: Agreement and reliability (in Matthews, DE. Farewell, VT: Using and understanding medical statistics, 4th ed., Karger, 2007, pp. 298–309.)

6．Eliasziw M, Young MA, *et al.*: Statistical methodology for the concurrent assessment of interrater and intrarater reliability: using goniometric measurements as an example. Physical Therapy. 1994;74(8):777–788.

7．柴田賢一，森嶋直人，宮原英夫：足関節底屈筋力測定の信頼性．愛知県理学療法学会雑誌．投稿中

8．Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–174.

図表

表1．検査日別に集計した足関節底屈筋力（Nm）の対象者別測定値，5回の測定の平均，標準偏差（SD）

1日目

ID	1回目	2回目	3回目	4回目	5回目	平均	SD
A	125.8	119.1	130.6	111.6	126.1	122.64	7.41
B	122.2	120.1	116.1	110.1	95.6	112.82	10.67
C	97.1	95.1	84.9	89.9	99.5	93.30	5.88
D	109.7	102.8	127.9	151.6	151.1	128.62	22.69
E	78.2	76.6	87.1	89.9	97.0	85.76	8.46
F	87.1	84.1	99.8	79.9	81.6	86.50	7.91
G	171.4	162.2	153.9	164.9	163.7	163.22	6.28
H	152.4	140.9	139.4	140.1	149.0	144.36	5.93
I	121.2	132.3	125.6	119.1	117.6	123.16	5.93
J	126.1	119.9	116.8	123.9	111.7	119.68	5.72
平均	119.12	115.31	118.21	118.10	119.29	118.006	1.60

2日目

ID	1回目	2回目	3回目	4回目	5回目	平均	SD
A	139.5	158.5	156.2	167.9	173.6	159.14	11.67
B	130.4	131.7	134.8	129.5	131.5	131.58	1.79
C	94.2	86.1	82.4	81.1	75.3	83.82	6.24
D	139.9	145.0	159.2	134.0	157.2	147.06	9.76
E	127.5	119.6	111.7	114.4	102.9	115.22	8.18
F	125.3	133.7	132.8	122.0	139.3	130.62	6.20
G	130.3	128.0	131.1	132.1	139.3	132.16	3.82
H	147.7	135.3	143.3	138.9	135.6	140.16	4.75
I	111.5	118.7	116.9	114.2	112.1	114.68	2.76
J	116.5	112.8	118.9	107.7	119.9	115.16	4.46
平均	126.28	126.94	128.73	124.18	128.67	126.96	1.69

3日目

ID	1回目	2回目	3回目	4回目	5回目	平均	SD
A	138.7	123.3	142.1	147.9	171.5	144.70	17.53
B	148.9	149.4	144.5	150.1	146.4	147.86	2.34
C	84.1	93.3	93.6	94.1	74.3	87.88	8.65
D	166.9	168.7	162.2	141.6	136.1	155.10	15.15
E	126.5	112.3	114.4	107.5	115.5	115.24	7.00
F	127.9	129.2	127.9	117.6	109.8	122.48	8.49
G	124.2	122.4	125.3	115.0	133.8	124.14	6.73
H	159.2	152.6	150.4	153.0	144.7	151.98	5.22
I	99.1	90.0	99.3	100.3	91.7	96.08	4.83
J	129.1	133.7	132.3	125.0	126.1	129.24	3.78
平均	130.46	127.49	129.2	125.21	124.99	127.47	2.41

表２．検査日別，被検者別に整理した測定値の構造

	検査日		
	第１日目	第j日目	第t日目
被検者番号	セッション内の測定値		セッション内の測定値
1	X_{111}, ・・・ , X_{11m}	・・・・	X_{1t1}, ・・・ , X_{1tm}
2	X_{211}, ・・・ , X_{21m}	・・・・	X_{2t1}, ・・・ , X_{2tm}
・	・ ・ ・	・	・ ・ ・
・	・ ・ ・	X_{ijk}	・ ・ ・
・	・ ・ ・	・	・ ・ ・
n	X_{n11}, ・・・ , X_{n1m}	・・・・	X_{nt1}, ・・・ , X_{ntm}

i＝1, 2, …, n;　　j＝1, 2, …, t;　　k＝1, 2, …, m

表３．繰り返しのある２元配置分散分析表
要因，自由度および平均平方とその期待値

変動要因	自由度	平均平方	期待値
被検者(A)間の変動	n－1	MSS	$\sigma^2+m\times\sigma^2_{AB}+m\times t\times\sigma^2_A$
測定日(B)間の変動	t－1	MSR	$\sigma^2+m\times\sigma^2_{AB}+m\times n\times\sigma^2_B$
交互作用(AB)の変動	(n－1)×(t－1)	MSSR	$\sigma^2+m\times\sigma^2_{AB}$
測定日内の変動	n×t×(m－1)	MSE	σ^2
全変動	n×t×m－1		

表４．１元配置分散分析表
要因，自由度および第i日目のデータの平均平方とその期待値

変動要因	自由度	平均平方	期待値
第i日の被検者(A)間の変動	n－1	MSS(i)	$\sigma^2+m\times\sigma^2_A$
第i測定日内の変動	(m－1)×n	MSE(i)	σ^2
全変動	n×m－1		