## 近年の実践研究で国際的に用いられる分析方法の例

#### 永原 隆

鹿屋体育大学スポーツ・武道実践科学系

連絡先: 〒891-2393 鹿児島県鹿屋市白水町1 鹿屋体育大学

E-mail: nagahara@nifs-k.ac.jp Tel: 0994-46-5231

# はじめに

近年、スポーツ科学に関する研究は盛んに行われているが、研究から得られた知見の有効性に関して、「試合に向けたアスリートのトレーニングをどう評価しているかについて過去と現在を比較した結論として、生理学に関する多数の知見を証明し、なぜ高いパフォーマンスを生じさせたのか説明することができるようになったかもしれないが、研究者が考えるほど研究成果によってコーチがアスリートを導く能力は向上していない」と指摘されることもあり(Foster 2016)、役に立つ実践研究の需要は国内外において高まっている。このような背景から、実践研究を標榜する国際的研究誌は、近年増加しており、2006年に創刊されたInternational Journal of Sports Physiology and Performance は、インパクトファクターが3.384(2017年)となっている(注1).この値は、スポーツ科学における親科学の主要な研究誌である Journal of Applied Physiology や Journal of Biomechanics のそれを上回っており、実践研究のスポーツ科学における価値は(インパクトファクターという観点では)高まっていると言える。これらの研究誌では、伝統的に広く用いられてきた分析方法(主に統計的方法)に加えて、新たな分析方法の使用が提案され、用いられている現状がある。そのような状況を鑑み、本稿では近年国際誌において実践研究に用いられている分析方法の例について紹介する。なお、ここでは分析方法を紹介するのであって、その利用を推奨し、伝統的な分析方法を忌避することを勧めるものではないことに留意いただきたい。

# Magnitude based inference (MBI) approach

実践研究に限らず,近年では不必要にサンプルサイズを大きくすることや,有意差が現れた統計分析法を選んで用いる「p-hacking」や「search for asterisks」が問題となり(Meyer et al. 2017),帰無仮説に基づく検定を使う研究の不誠実さへの指摘が散見される.このような背景から,帰無仮説に基づく(p値を使う)検定を禁止する Basic and Applied Social Psychology(Trafimow and Marks 2015)や Sport Performance & Science Reports(注 2)などの研究誌もある.帰無仮説に基づく検定を用いない方法として,近年実践研究で多くみられるものに Magnitude-based inference (MBI) approach がある(Batterham and Hopkins 2006). MBI は,2006年にスポーツ科学向けに初めて体系化された形で提案され,現在まで数多くの実践研究で用いられている(Ayala et al. 2017, Bezodis et al. 2017, Higham et al. 2014, Nagahara et al. 2018, Wild et al. 2018).

MBI については、反対する立場から「MBI は明らかに Type I errors を生じさせる危険があり、使うべきではない」(Welsh and Knight 2015a)といった指摘がある一方、その使用を推奨する立場から「帰無仮説に基づく統計分析より優れているのは事実」(Hopkins and Batterham 2015)と反論があるなど、その使用の是非については議論が続いている(Batterham et al. 2018, Borg et al. 2018, Nevill et al.

2018, Sainani 2018, Welsh and Knight 2015b, Young 2015). しかし、スポーツ科学や臨床医学などの実践的/医学的な効果が統計的な有意性より重要性を持つ分野で広く使われるようになっている現状がある(Buchheit 2017). 帰無仮説に基づく統計の限界として、「p値(研究の結論)が、サンプルサイズによって変化すること」、「有意であることが、効果(有効性)の大きさと関係しないこと(200 名の選手が0.01%変化した場合でも有意になる)」、「有意な結果のみが投稿される傾向があること」が挙げられている(Buchheit 2016, 2017). このような帰無仮説に基づく研究の限界に対して、MBI は「サンプルサイズに影響されず、変化の大きさを適切に評価できること(p値が有意でない場合でも、効果の大きさを評価できる。意味のある差がないのか(信頼区間が Smallest worthwhile change (SWC)の範囲内なのか)、差があるかはっきりしないのか(信頼区間が SWC の範囲を超えているのか)を区別して考えることができる)」、「個人内の変化を評価することにも使えること(個人の測定値の変化は、いくつかの方法(z スコア、standard difference score など)で評価できるが、標準誤差・SWC を設定することで、MBI はさらに尤度を評価できる)」、「データのビジュアル化を改善できること」が利点として挙げられている(Buchheit 2017).

MBI の基本的なコンセプトは、設定した SWC (例えば、パフォーマンス変数の 1%) に対して、パフォー マンス変数の変化量の 95%信頼区間の範囲がどこに位置しているかで評価する(Batterham and Hopkins 2006) (図 1). その際, 閾値を設定し, 変化について「変化が明確ではない, ポジティブな変化 があった, ネガティブな変化があった, 変化がなかった」といった解釈をする. 帰無仮説に基づく分析方 法の「全か無か」ではなく、「どの程度有益であるのか」を評価するのが特徴である. 実際に MBI を使う 場合の例として, 1)効果量(Cohen's d)とその信頼区間(95%)を算出し, 2)SWC(効果量 0.2)を決め, 3) Percentage likelihood を求めるのが手順である. SWC は, 意味のある最小の変化量として任意に決 定する. Percentage likelihood は、効果がポジティブ、ネガティブ、取るに足りないものなのか、それらの 推定割合(%)を示すものである. Percentage likelihood の閾値は, most unlikely (<0.5%), very unlikely (<5%), unlikely(<25%), possibly(<75%), likely(<95%), very likely(<99.5%), most likely(>99.5%)と設 定されることが多い(Hopkins 2006). また, 結果を視覚的に表現するために, 平均値, 95%信頼区間, Percentage likelihood などを図に示すことが一般的である(図 2). 先に述べたように、MBI では帰無仮 説に基づく統計とは異なり, 研究者(実践者)が事前に重要な影響を及ぼす最小量(SWC)を定義し, それを尤度の決定に使う必要がある. 一般的に SWC は個人では変動係数(coefficient of variation, CV)の33%,グループを対象とする場合は、SDの20%が用いられる. ただし、SWC は状況に応じて適切 に設定すべきとされている. 例として, サッカーで相手に競り(走り)勝つ(約 30cm の差を生み出す)た めには、20m 走のタイムが 1%改善すれば良く、その値が実践的意味を持つものになる(Buchheit 2017).

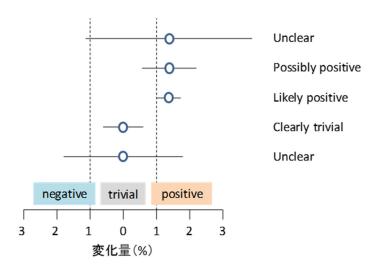


図 1 MBI を用いた測定値変化を評価した例. 白抜きの丸は変化の大きさ、横棒は 95%信頼区間を示す

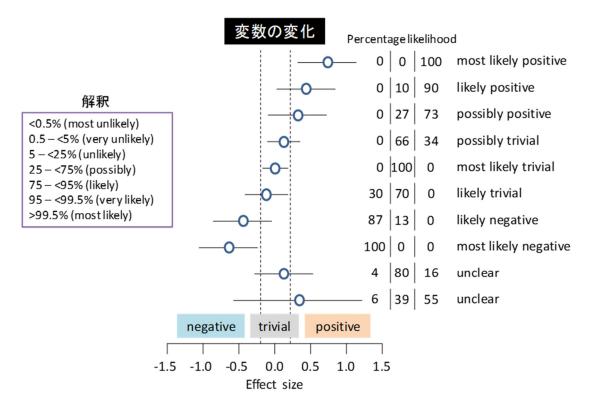


図 2 MBI を用いた分析における結果とその解釈の例. 白抜きの丸は変化の大きさ, 横棒は 95%信頼区間を示す. 右側の数字は, 95%信頼区間の negative, trivial, positive の範囲における割合 (percentage likelihood)を示したものである.

最後に MBI を用いた分析について、具体例を示したい。Bezodis et al. (2017)は、チームスポーツの選手に 1) 普段通り、2) external focus、3) internal focus という 3 つ異なる意識の条件で 10m 走を行わせ、条件間のパフォーマンスやバイオメカニクス変数の差について、MBI を用いて明らかにしている (SWC は、効果量(Cohen's d) 0.2 に設定、解釈は上記の閾値に同じ)。図 3 は、10m 走のタイム差について MBI を使って分析した結果を示したものであり、external focus、internal focus と普段通りの間の

差は、それぞれ効果量 0.5 前後で、信頼区間が正負の SWC を上回ることなく、positive な percentage likelihood がそれぞれ 98%、96%であり(3 条件なので効果量の 97%信頼区間)、これらの条件間の差は very likely と解釈され、「external focus, internal focus よりも普段通り走ったほうがタイムが良さそうだ」 との結論を得ている。一方、external focus と internal focus の間の差における効果量はゼロに近いが、信頼区間は正負の SWC を上回っているため、両者の差は unclear と解釈され、「external focus と internal focus の間で疾走タイムの差ははっきりしない(差がないという結論ではない事に注意).

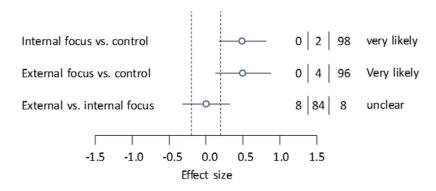


図 3 MBI を用いた分析結果の例 (Bezodis et al. 2017). 3 つの条件間における 10m 走タイムの比較. 白抜きの丸は変化の大きさ、横棒は 97%信頼区間を示す. 右側の数字は、97%信頼区間の negative, trivial, positive の範囲における割合 (percentage likelihood)を示したものである.

図 4 は、個人内のトレーニング (12 km/h ron = 2 ron) における心拍数の変化を長期にわたって評価したものである (Buchheit 2017). この研究では、あらかじめ設定した SWC に対して、心拍数と typical error がどこに位置するかで、その心拍数が低いのか高いのか、その蓋然性がどの程度であるかを評価している。このように、平均値や SD, CV がない場合でも、SWC を任意に設定し、個人内のパフォーマンスを評価・解釈できる点が MBI の特徴である。

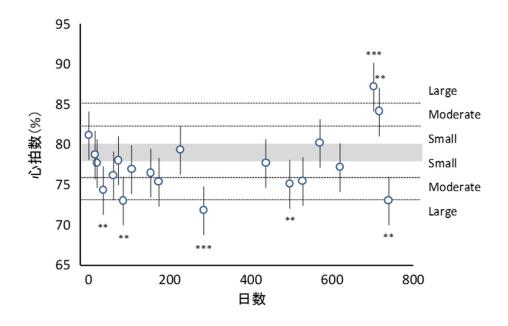


図 4 MBI を用いた分析結果の例 (Buchheit 2017). プロサッカー選手の 12km/h でのランニングにおける最大化心拍数の 2 シーズンにわたる推移 (最大心拍に対する比[%]). 灰色の背景は, trivial (1%) な変化の範囲を示す. エラーバーは, typical error (3%), アスタリスクの数は, 変化が意味のあるものである見込みをそれぞれ示す. \*\*, likely; \*\*\*, very likely. 変化の度合いは SWC の何倍であるかによって定義. 1-3 倍, small; 3-6 倍, moderate; 6 倍以上, large.

### ブートストラップ法

ブートストラップ法は、長期にわたるパフォーマンスのモニタリング研究を中心に用いられている (Bezodis et al. 2018, Dixon et al. 2014, Hellard et al. 2006, Joch et al. 2018, Lenohoff et al. 1999, Pelechrinis et al. 2016, Salo et al. 2010, Salo et al. 2011). ブートストラップ法は、母集団の推定量(分散など)の性質を近似分布にしたがって標本化したときの性質を計算することで推定する分析方法であり、新しいデータセットを取得して統計値を求める. 例えば、任意の回数(1000回, 10000回など)リサンプリングを繰り返せば、それらの回数分のデータの平均値と標準偏差、信頼区間を得ることが出来るため、(特に少ないサンプルで)統計量を推定する場合に役立つ(Efron and Tibshirani 1993). ここでは、具体的に例を示して、どのような研究で用いられているか紹介する.

Salo et al. (2011) は、世界トップレベルのスプリンターにおける数年にわたる 100m 走のパフォーマンス変化が、ステップ長とステップ頻度のどちらに強く影響を受けているかについて、ブートストラップ法を用いて個々の選手のタイム-ステップ頻度関係、タイム-ステップ長関係の差の 90%信頼区間を求めることで検討している。その過程では、個々の選手 100m 走を複数レース分析し(11-27回)、得られた 100m 走のタイム、ステップ長、ステップ頻度を対数変換後、個々の選手のサンプルを 10000 回リサンプリングしている。その後、1 回ごとにタイムとステップ長、ステップ頻度それぞれの相関係数を計算し、両者の差(ステップ頻度ーステップ長)を求め、蓄積している。そして、その相関係数の差の 90%信頼区間を得て、設定した閾値(相関係数 0.1)に対して信頼区間がどこに位置するかで、ステップ頻度型かステップ長型かを判定している。

一方、Bezodis et al. (2018)は、冬季トレーニング中にスプリント走パフォーマンス(速度)とそれを構成するステップ長、ステップ頻度がどのように変化したかをモニタリングした研究でブートストラップ法を用いている。この研究では、取得したデータ数が個人ごと、セッションごとに異なることが問題であったが(3-27回)、ブートストラップ法によってデータをリサンプリングすることで、各セッションのデータ数を1000回として分析している。その過程では、リサンプリング後の速度、ステップ長、ステップ頻度について個人内で反復測定分散分析と多重比較検定(Dann-Sidak 補正)を行い、個人ごとにセッション間の差の有無を評価している。また、セッション間の差の効果量を求め、Hopkins et al. (2009)の基準に従って、その大きさについて解釈している。このような方法で、個人ごとの速度、ステップ長、ステップ頻度の変化を評価し、速度の変化に対して、ステップ長、ステップ頻度どちらの変化を伴っているかを検討することで、トレーニングにおいて注目すべき変数(ステップ長のアステップ頻度)を明らかにしている(図5)。

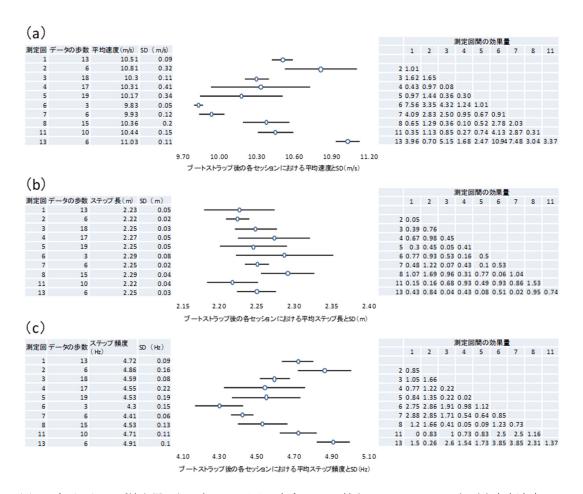


図 5 ブートストラップ法を用いた 1 名のモニタリングデータの比較 (Bezodis et al. 2018). (a) 疾走速度, (b) ストライド, (c) ピッチ. 右端の表は, 測定回間の差の効果量(Cohen's d)を表す.

両研究のポイントは、縦断的変化を集団ではなく、個人ごとに評価していることである。この方法を用いることで、「対象者のうち何名は~」といった主張をより高い蓋然性を伴って行うことができる。また、Bezodis et al. (2018)では、4名のみを対象者としており、少ないサンプルサイズでも実践的に有用であ

れば、論文として受理され、出版される例となっている。さらに、このブートストラップと MBI を組み合わせて、個人にそのトレーニング(介入)などがどの程度有効なのか示す方法も提案されている(注 3).

#### Tau-U

Tau-U は、2011 年に Parker らによって提案された single case design のための効果量 (-1-1) の指標 (複数) である (Parker et al. 2011). Tau-U は、Mann-Whitney U 検定とケンドールの順位相関係数 (ケンドールの Tau) から導かれた指標であり、例えば図 6 に示したような AB デザインのベースラインにおける値の変化の特性を考慮し、介入の効果を評価する点が特徴である (ベースラインの傾向を補正しない場合、結果は Mann-Whitney U 検定と同等). Tau-U は、個人 <math>(少人数) のトレーニング (介入) にともなう測定値の変化を評価する場合などに用いることができ  $(Harnish\ et\ al.\ 2018$ 、Lee and Cherney 2018、Stylianou et a. 2016)、スポーツパフォーマンスを対象とした実践研究にも応用できると考えられる. また、ウェブ上で値を求めることができるのも利点といえる  $(注\ 4)$ .

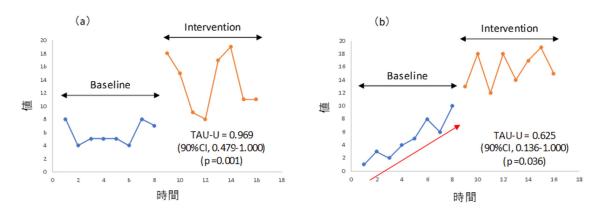


図 6 AB デザインにおける Tau-U の結果の例. (a) はベースライン, 介入の傾向を補正する必要がなかった場合の例, (b) はベースラインの傾向を補正した場合の例.

ここでは、Tau-Uを使用した研究の例として、脳卒中後の失語症に対する有酸素運動やストレッチの補助としての効果を個人ごとに検証した研究を紹介したい(Harnish et al. 2018).この研究では、失語症治療のみと、それに続く有酸素運動(5名)あるいはストレッチ(2名)を伴った治療について、写真を使ったネーミングテスト結果に関して、Tau-Uを用いて評価している。その結果、有酸素運動では5名中3名により大きな治療の効果があったこと、ストレッチでは2名中1名により大きな治療の効果があったことを報告している。また、個々の患者の治療効果を評価したTau-Uを統合した結果から導かれたグループレベルでの比較でも有酸素運動を伴う治療の効果が大きかったことを指摘している。このように、Tau-Uは、少人数を対象とした介入の効果検証といった場面で用いられている。

### おわりに

本稿では,近年国際的研究誌で用いられている分析方法について紹介した.紹介した分析方法以外にも,ベイズ統計を用いた研究など(Carvalho et al. 2017, Hecksteden et al. 2017, Klemish et al. 2018),これまで一般的ではなかった分析方法でスポーツの実践研究が行われるようになっている.一

方, 実践的なスポーツ科学研究誌に限らず, 伝統的に用いられている帰無仮説に基づく分析方法を 用いる場合でも,近年では実際の p 値を示すことが求められ(American Psychological Association 2009), 「p < 0.05」とするのではなく, 「p = 0.023」などと表記することが推奨されている. つまり, 有意で あったか否かだけでなく、実際の確率を示すことが求められている. また、p 値(有意確率)に加えて効 果量(Cohen's d, r,  $\omega^2$ など)を示すことも一般的になっている(American Psychological Association 2009). これらの変化の潮流は, 既存の分析方法の不誠実さと役に立つ研究成果の追求が背景にある だろう。そのような流れのなかで、研究誌としてのスポーツパフォーマンス研究も前衛的な分析方法を 積極的に受け入れ, 実践に役立つ成果を公表する土俵を作る必要があるかもしれない. 実践研究に 限らず, ネガティブデータ(差, 効果, 関係がないと判断された結果や研究)は闇に葬られ, ポジティブ データのみが投稿され、論文として出版されることが多い.このことは、バイアスとして捉えられている. しかし,スポーツ活動においては,パフォーマンス向上を目指して試行錯誤を繰り返しており,成功事 例を上回る失敗事例が存在することは想像に難くない. また, 競技における技術の習得には, 成功す る手段, 方法を模索するより, 失敗する手段, 方法を確認し, 消去していく作業の方が, 効率が良いよう にも感じられ、ネガティブデータに価値がないとは言えないだろう. 近年では、出版時に査読を行わず (内容が趣旨に合致し、ガイドラインに従っていれば全て出版する)、出版後にその研究論文について 議論するスタイルのスポーツパフォーマンスに特化した国際研究誌(注 5)も発刊されており,事例や少 ないサンプルサイズの研究,追試が奨励されており、ネガティブデータでも積極的に投稿することが求 められている. 本稿で紹介した方法を用いた研究が、どの程度国内の研究誌で受け入れられるか定か ではないが、スポーツパフォーマンス研究に限らず、査読を行わない研究誌やネガティブデータのみを 出版する研究誌など、大胆な発想で実践にある知を記録させる取り組みがスポーツにおける実践研究 に必要かもしれない.

- (注 1) https://journals.humankinetics.com/page/about/ijspp (2019年1月22日閲覧)
- (注 2) https://sportperfsci.com/authors-guidelines/ (2019年1月22日閲覧)
- (注 3) http://complementarytraining.net/bmbp/ (2019年1月31日閲覧)
- (注 4) http://www.singlecaseresearch.org/calculators/tau-u (2019年1月29日閲覧)
- (注 5) Sport Performance & Science Reports

出版後に読者からコメントを受け付け、著者が答える. それに伴い、著者が望めば出版後の論文を修正することができる. 修正前後の全ての論文、読者からのコメント、それへの返答が保存され、参照できる.

### 文献

- · American Psychological Association (2010). Publication manual of the American Psychological Association. 6th ed. Washington, DC, American Psychological Association.
- · Ayala F, Calderon-Lopez A, Delgado-Gosalbez JC, Parra-Sanchez S, Pomares-Noguera C, Hernandez-Sanchez S, Lopez-Valenciano A, De SCM (2017) Acute Effects

- of Three Neuromuscular Warm-Up Strategies on Several Physical Performance Measures in Football Players. PLoS One 12:e0169660.
- · Batterham AM, Hopkins WG (2006) Making meaningful inferences about magnitudes. Int J Sports Physiol Perform 1:50-57.
- · Batterham AM, Hopkins WG (2015) The case for magnitude-based inference. Med Sci Sports Exerc 47:885.
- · Batterham AM, Hopkins WG (2018) The Problems with "The Problem with 'Magnitude-based Inference'". Med Sci Sports Exerc [epub ahead of print]
- · Bezodis IN, Kerwin DG, Cooper SM, Salo AIT (2018) Sprint running performance and technique changes in athletes during periodized training: An elite training group case study. Int J Sports Physiol Perform 13:755-762.
- · Bezodis NE, North JS, Razavet JL (2017) Alterations to the orientation of the ground reaction force vector affect sprint acceleration performance in team sports athletes. J Sports Sci 35:1817-1824.
- · Borg DN, Minett GM, Stewart IB, Drovandi CC (2018) Bayesian Methods Might Solve the Problems with Magnitude-based Inference. Med Sci Sports Exerc 50:2609-2610.
- · Buchheit M (2016) Chasing the 0.2. Int J Sports Physiol Perform 11:417-418.
- Buchheit M (2017) Want to see my report, coach? Aspeter Sports Medicine Journal 6:36-43.
- · Carvalho HM, Goncalves CE, Grosgeorge B, Paes RR (2017) Validity and usefulness of the Line Drill test for adolescent basketball players: a Bayesian multilevel analysis. Res Sports Med 25:333-344.
- Dixon PC, Stebbins J, Theologis T, Zavatsky AB (2014) Ground reaction forces and lower-limb joint kinetics of turning gait in typically developing children. J Biomech 47:3726-3733.
- · Efron B, Tibshirani R (1993) An introduction to the bootstrap. Monographs on Statistics & Applied Probability. New York: Chapman & Hall.
- · Foster C (2016) Back to the future. Int J Sports Physiol Perform 11:281-282.
- · Harnish SM, Rodriguez AD, Blackett DS, Gregory C, Seeds L, Boatright JH, Crosson B (2018) Aerobic Exercise as an Adjuvant to Aphasia Therapy: Theory, Preliminary Findings, and Future Directions. Clin Ther 40:35-48.e6.
- · Hecksteden A, Pitsch W, Julian R, Pfeiffer M, Kellmann M, Ferrauti A, Meyer T (2017) A New Method to Individualize Monitoring of Muscle Recovery in Athletes. Int J Sports Physiol Perform 12:1137-1142.
- · Hellard P, Avalos M, Lacoste L, Barale F, Chatard JC, Millet GP (2006) Assessing the limitations of the Banister model in monitoring training. J Sports Sci 24:509-520.
- · Higham DG, Hopkins WG, Pyne DB, Anson JM (2014) Performance indicators related

to points scoring and winning in international rugby sevens. J Sports Sci Med 13:358-364.

- · Hopkins WG (2006) Spreadsheets for analysis of controlled trials with adjustment for a predictor. Sportscience 10:46-50.
- · Hopkins WG, Marshall SW, Batterham AM, Hanin J (2009) Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc 41:3-13.
- · Joch M, Dohring FR, Maurer LK, Muller H (2018) Inference statistical analysis of continuous data based on confidence bands-Traditional and new approaches. Behav Res Methods [epub ahead of print]
- · Klemish D, Ramger B, Vittetoe K, Reiter JP, Tokdar ST, Appelbaum LG (2018) Visual abilities distinguish pitchers from hitters in professional baseball. J Sports Sci 36:171-179.
- · Lee JB, Cherney LR (2018) Tau-U: A Quantitative Approach for Analysis of Single-Case Experimental Data in Aphasia. Am J Speech Lang Pathol 27:495-503.
- · Lenhoff MW, Santner TJ, Otis JC, Peterson MG, Williams BJ, Backus SI: Bootstrap prediction and confidence bands: a superior statistical method for analysis of gait data.. Gait Posture 9:10-17,1999
- · Meyer KE, van Witteloostuijn A, Beugelsdijk S (2017) What's in a p? Reassessing best practices for conducting and reporting hypothesis-testing research. J Int Bus Stud 48:535-551.
- · Nagahara R, Mizutani M, Matsuo A, Kanehisa H, Fukunaga T (2018) Step-to-step spatiotemporal variables and ground reaction forces of intra-individual fastest sprinting in a single session. J Sports Sci 36:1392-1401.
- · Nevill AM, Williams AM, Boreham C, Wallace ES, Davison GW, Abt G, Lane AM, Winter EM EDITORIAL BOARD (2018) Can we trust "Magnitude-based inference"? J Sports Sci 36:2769-2770.
- Parker RI, Vannest KJ, Davis JL, Sauber SB (2011) Combining nonoverlap and trend for single-case research: Tau-U. Behav Ther 42:284-299.
- · Pelechrinis K, Papalexakis E (2016) The Anatomy of American Football: Evidence from 7 Years of NFL Game Data. PLoS One 11:e0168716.
- Salo AI, Bezodis IN, Batterham AM, Kerwin DG (2011) Elite sprinting: are athletes individually step-frequency or step-length reliant? Med Sci Sports Exerc 43:1055-1062.
- · Salo PK, Hakkinen AH, Kautiainen H, Ylinen JJ (2010) Effect of neck strength training on health-related quality of life in females with chronic neck pain: a randomized controlled 1-year follow-up study. Health Qual Life Outcomes 8:48.
- · Sainani KL (2018) The Problem with "Magnitude-based Inference". Med Sci Sports

Exerc 50:2166-2176.

- · Stylianou M, Kulinna PH, van der Mars H, Mahar MT, Adams MA, Amazeen E (2016) Before-school running/walking club: Effects on student on-task behavior. Prev Med Rep 3:196-202.
- · Trafimow D, Marks M (2015) Editorial. Basic Appl Soc Phych 37:1-2.
- · Wild JJ, Bezodis IN, North JS, Bezodis NE (2018) Differences in step characteristics and linear kinematics between rugby players and sprinters during initial sprint acceleration. Eur J Sport Sci 18:1327-1337.
- · Welsh AH, Knight EJ (2015a) "Magnitude-based inference": a statistical review. Med Sci Sports Exerc 47:874-884.
- · Welsh AH, Knight EJ (2015b) Response. Med Sci Sports Exerc 47:886.
- · Young AJ (2015) Magnitude-based inference: progressive approach or flawed statistic? Med Sci Sports Exerc 47:873.