Monitoring fatigue status in elite team sport athletes: Implications for practice

Brief Review

RT. Thorpe¹,², G. Atkinson³, B. Drust², W. Gregson²

¹ Medicine and Science Department, Manchester United, Manchester, UK
² Football Exchange, Research Institute for Sport and Exercise Sciences, Liverpool
³ John Moores University, UK

Corresponding Author:
Robin T. Thorpe
Manchester United, Aon Training Complex, Birch Road off Isherwood Road, Carrington, Manchester M31 4BH
E-mail: robin.thorpe@manutd.co.uk

Key Words-Training, Performance, Wellness, Recovery, Injury, Illness

Word count: 4315
Abstract word count: 223
Number of figures: 0
Number of tables: 0
Abstract

The increase in competition demands in elite team sports over recent years has prompted much attention from researchers and practitioners into the monitoring of adaptation and fatigue in athletes. Monitoring of fatigue and gaining an understanding of athlete status may also provide insights and beneficial information pertaining to player availability, injury and illness risk. Traditional methods used to quantify recovery and fatigue in team sports such as maximal physical performance assessments may not be feasible in order to detect variations in fatigue status throughout competitive periods. The implementation of more quick, simple and non-exhaustive tests such as athlete self-report measures (ASRM), autonomic nervous system (ANS) response via heart rate derived indices and to a lesser extent jump protocols may serve as promising tools to quantify and establish fatigue status in elite team sport athletes. The robust rationalization and precise detection of a meaningful fluctuation in these measures are of paramount importance for practitioners working alongside athletes and coaches on a daily basis. There are various methods for arriving at a minimal clinically important difference (MCID), but these have been rarely adopted by sports scientists and practitioners. The implementation of appropriate, reliable and sensitive measures of fatigue can provide important information to key stakeholders within team sport environments. Future research is required to investigate the sensitivity of these tools to fundamental indicators such as performance, injury and illness.

Introduction

Elite team sport athletes, particularly those in the professional football codes, are exposed to high competition loads, particularly in recent years. These high loads reflect a number of factors, including an increased frequency of domestic competitions, particularly for higher-level athletes, as well as a higher intensity of competition due to enhanced player preparation strategies. Higher loads may also result from the increased demands of international competition during both the domestic season and the off-season period.

An increased availability of athletes for selection, as a result of a reduction in injuries, substantially increases a team’s chance of success. Therefore, injury prevention strategies are fundamental to the work of the athlete’s support team. Routine modifications in training load (frequency, duration, intensity) occur during the training cycle and these subsequently increase or decrease fatigue. Management of fatigue is important in mediating adaptation to training and ensuring the athlete is prepared for competition, as well as for reducing the athletes susceptibility to non-functional overreaching, injury and illness.

The importance of managing athlete fatigue has led to an increase in interest in monitoring athlete loads, particularly in terms of the measures which may offer insights into whether the athlete is adapting positively or negatively to the collective stresses of training and competition. In the present review, we will consider published research concerned with the monitoring of fatigue status in team sport athletes. Information derived from other sports will be examined where evidence in team sport settings is not available. While there are studies in which various proposed
moderators and mediators have been found to be statistically significantly associated with fatigue status, in our paper we also highlight the various measurement issues and practical considerations which should be considered by those responsible for the development and implementation of player monitoring systems in the field. We focus, especially, on the neglected topic of all the different approaches to selecting a minimal worthwhile change in fatigue status. It is hoped that our overview will provide the basis for the development of a framework for evaluating fatigue status in team sports and provide some guidance for future investigators.

It is not our intention to comprehensively describe the available information on the etiology of fatigue or the scientific basis of common recovery intervention practices. Such information can be found in several excellent recent reviews. For the purpose of this review, and to align with previous reviews in this area, fatigue will be defined as “an inability to complete a task that was once achievable within a recent time frame”.

**Methods for Monitoring Fatigue**

Training load reflects the internal and external loads imposed upon the athlete. External load relates to work completed by the athlete independent of his or her internal characteristics and is important for understanding the capabilities and capacities of the athlete. The internal load, or the relative physiological strain resulting from the external training factors, is also crucial to determining both the stress imposed and subsequent adaptation to training. A combination of both the external and internal load is therefore important for training since the uncoupling or divergence of external and internal loads may differentiate between a non-fatigued and a fatigued athlete. This approach is particularly relevant in “closed loop” sports like cycling where the performance outcome is time, and the power produced by the rider is known to have a relatively precise association with the performance time. Under such conditions, the internal load needed to sustain a certain external load (power output) can provide important information regarding the athlete's fatigue status.

In contrast to “closed loop” sports, the ability to relate external and internal loads in “open loop” sports like team sports is difficult due to the inherent variability in physical performance during sport-specific training drills and match-play. As a consequence, attempts to monitoring the fatigue status of team sport athletes have largely focused on the assessment of internal and external load measures under resting conditions and/or during submaximal exercise assessments on the morning prior to training. Within the confines of this approach, a valid indicator of fatigue in team sports should be sensitive to training load and their response to acute exercise should be distinguishable from chronic changes in adaptation. Prospective tools should also be non-invasive, quick and easy to administer and limit any additional loading on the athlete. This is particularly important in football codes, where competition occurs on a weekly basis and in some instances on 2-3 occasions a week meaning that players are required to peak with limited recovery between matches.
Athlete self-report measures

Recent surveys on fatigue monitoring in high performance sport demonstrate that athlete self-report measures (ASRM) are used extensively for assessing the overall well-being of team sport athletes. A plethora of ASRM currently exist including the POMS, DALDA, TQR and REST-Q, which have been extensively documented in the literature. However, many of these are often extensive and time-consuming to complete preventing their use on a daily basis with large numbers of team athletes. Many team sports therefore often adopt shorter, customised questionnaires which can be administered on daily basis. A recent review highlighted that ASRM demonstrate greater sensitivity to acute and chronic training loads compared to commonly used objective measures. In team sports, for example contemporary Australian Football League (AFL) and English Premier League (EPL) research has shown custom psychometric scales to be sensitive to daily, within-weekly and seasonal changes in training load. Indeed, daily ASRM (fatigue, sleep quality, stress, mood and muscle soreness) were significantly correlated with daily training load in a pre-season camp and competitive period in AFL and EPL players respectively. Similarly, ASRM were sensitive to changes in training load during typical weeks in AFL and EPL players across the course of the season. Further importance of ASRM and relationship with injury/illness has been observed in Rugby League, in this study fluctuations in ASRM between macrocycles were shown to provide useful insights into possible illness risk in players. Further work is required to examine the relationships between ASRM and injury/illness risk in team sport athletes.

Autonomic Nervous System (ANS)

The ANS is interlinked with many other physiological systems, significant attention in the literature has therefore centered upon the use of indicators of ANS functioning for determining an athletes overall adaptation/fatigue status. To date this has largely stemmed from studies examining the sensitivity of heart rate (HR) derived indices including resting heart rate (RHR), exercising heart rate (HRex) and heart rate variability (HRV) and heart rate recovery (HRR) to fluctuations in training and competition load.

Submaximal heart rate (HRex) Decreases in HR during standardised exercise bouts have traditionally been associated with increases in aerobic fitness. However, the majority of data available has reported inconsistent results in non-team sport athletes. Heart rate during intensified training and during varying intensities showed significant reductions in overreached triathletes. Le Meur and colleagues suggested a hyper-activation of the parasympathetic nervous system via central, cardiac and/or periphery mechanisms. Recent observations in AFL have also reported reductions in heart rate in response to training during pre-season, although, the authors concluded this was more possibly due to the effects of training/environmental induced changes in plasma volume than acute changes in fitness or fatigue. Contrary to these reports, exercising heart rate in EPL players failed to fluctuate in response to within week changes in training and match load over the course of a season. The use of HRex in healthy athletes to predict negative effects in performance or fatigue should be treated with caution and interpreted together with other potential measures of fatigue such as ASRM.
Heart rate variability (HRV) Vagal-related time domain parameters of HRV have recently received greater attention than more traditional spectral analyses due to their superior reliability and assessment capture over short periods of time. Sensitivity to changes in training load and performance has mainly been observed in non-team sports. Generally, HRV is reduced (sympathetic dominance) in the immediate days following intense exercise, however, results from endurance sports have shown inconsistent results. Little evidence currently exists with regards to its sensitivity to fluctuations in training and competition load in team sports. In AFL players undertaking pre-season training in the heat, a vagal-related HRV parameter (SD1) was largely and statistically significantly correlated ($r=0.5$) to daily RPE-TL. However, these unexpected changes in parasympathetic activity may have been partly mediated through thermoregulatory mechanisms associated with alterations in plasma volume. In elite soccer, HRV (Ln rMSSD) appeared to decrease ($r=-0.2$), albeit, transiently in response to high-speed running distance. Contrastingly, in the same population, HRV did not change across a standard in-season training week. Interestingly, data derived from endurance sports have suggested that the sensitivity of HRV to training and competition may be improved when data is averaged over a week or using 7-day rolling averages compared to the use of single data points due to the high day-to-day variation in these indices. However, undertaking such measures may prove difficult with the large volume of athletes engaged in team sports. Future research is needed to determine whether such approaches enhance the suitability of these measures for use in team sport populations.

Heart rate recovery (HRR) Post exercise HRR reflects general haemodynamic adjustments in relation to body position, blood pressure regulation and metaboreflex activity, which partly drives sympathetic withdrawal and para-sympathetic reactivation. Recent findings in endurance sports have shown that HRR may serve as a sensitive marker of acute training load alteration, although this association has yet to be seen in team sports. HRR did not fluctuate in response to daily and within week training load variability in EPL players. Data from physically active men and women have shown a delay in HRR following increases in training load. More recently, non-functionally overreached elite triathletes showed a faster (8 beats per min) HRR compared to elite triathlete controls following the same training program. It appears that HRR is responsive to both acute and chronic changes in training load, however, the exact direction of this change and whether HRR can detect fatigue status remains unclear and should be interpreted alongside training status and with caution.

Physical Performance

A variety of maximal performance assessments (sprints, repeated sprints, jumps and maximal voluntary contractions) have been used in attempt to quantify the rate of recovery of performance in the hours and days following training and competition in team sports. Whilst these types of assessment provide important information, the application of physical performance tests which are exhaustive in nature and time consuming to deliver means they are often unsuitable for use in team sport environments. Quick, efficient and without additional load represent the only feasible maximal performance assessments applicable for team sport players.
Neuromuscular function

Various jump protocols including squat jump (SJ) and countermovement jump (CMJ) have been adopted to examine the recovery of neuromuscular function following competition with significant decreases for up to 72-hours routinely reported. However, less attention has focused on examining their sensitivity to changes in training load. CMJ was not sensitive as a measure of neuromuscular status in EPL players when analysed alongside daily fluctuations in training load, furthermore, data derived from elite Rugby Sevens and adolescent soccer players revealed no change in countermovement jump height or correlation to training load during a taper and across a training period respectively. The use of jump height per se as a global indicator of neuromuscular function may lack the sensitivity to detect changes in training load in previous studies. Moreover, CMJ height alone may mask alternative neuromuscular measures and their sensitivity to alterations in load. Reductions in 18 different neuromuscular variables were found following a high-intensity fatiguing protocol in college-level team sport athletes. Neuromuscular parameters (eccentric, concentric, and total duration, time to peak force/power, flight time:contraction time ratio) derived from CMJ were deemed suitable for neuromuscular fatigue detection. In AFL, variations in force-time parameters (flight time:contraction time ratio) were observed over the course of a season, indicating sensitivity to increases in load over time. Future research is required to investigate whether alternative measures derived from CMJ are sensitive to changes in training load in elite team sport athletes.

Joint range of motion (JROM)/ Flexibility

Simple clinical assessments of JROM have been typically performed on a one-off basis as part of a pre-season screening battery in elite team sport. However, there is a lack of data reporting JROM responses to training and match load. Indeed, the assessment of JROM more regularly during competitive periods may provide greater information pertaining to structural fatigue and potential injury risk compared to a single pre-season assessment. In elite soccer players, knee range of motion was reduced until 48-hours post-match. Similarly, Mohr and colleagues (2015) found knee joint range of motion declined 7% at both 24- and 48-hours post-match. Moreover, structural assessments quantifying hip/groin extensibility have shown good reliability and validity following match-play in youth soccer players. Indeed, a reduction of more than 12.5% in adductor (bent knee fall out test) range 14-hours post-match was deemed a meaningful change in youth soccer players. The simple and quick nature of JROM assessments evaluating key anatomical regions may provide a greater understanding of structural status and potential injury risk. Future research is required, to examine the time-course of recovery for JROM measures post-match and their sensitivity to changes in load in team sport athletes.

Biochemical/Hormonal/Immunological

A large amount of research has examined a range of biochemical, hormonal and immunological responses to team sport competition. It is beyond the scope of this article to review the collective literature surrounding the responses of such measures in team sports, however, no definitive marker has been derived for examining the fatigue status of athletes. Furthermore, the associated costs and in some instances time
consuming nature of their analyses, often means many of these measures are impractical for use in the team sport environment.

A variety of markers have been used in an attempt to examine potential levels of muscle damage in athletes. Creatine kinase (CK) increases immediately post-match in soccer and rugby and peaks between 24-48-hours with a return to baseline values observed from 48-120-hours. Although widely used, questions remain regarding the exact mechanism of activity following exercise and its relationship with muscle function recovery. IL-6 is produced in larger amounts than any other cytokine prompting its use as a global measure of inflammation. IL-6 peaks immediately following the cessation exercise and then rapidly returns to baseline values after 24-hours. C-reactive protein (CRP) and uric acid have been found to be a more sensitive marker of inflammation following soccer match-play. Indeed, increases of up ~50%, 48-hours post-match has been observed in elite soccer players. Furthermore in a similar study in elite soccer players, uric acid peaked 72-hours following a match. Adrenal hormones cortisol and testosterone have been shown to increase up to 48-hours following competition and up to 50% post-competition in team sports respectively. Salivary immunoglobulin A (S-IgA) has become a popular means to assess mucosal immunity in athletes following via the use of real time lateral flow devices. In EPL players S-IgA showed reductions during a taper phase and a period of international competition, Little longitudinal data particularly around competition and training phases currently exists in team sports. The impractical nature and cost of individual samples may explain the limited data assessing biochemical, hormonal and immunological measures over extended training and competition periods in team sports.

**Measurement Considerations**

In a recent BJSM editorial, McCall et al. (2016) described the concept of “working fast - working slow”, whereby the researcher undertakes robust and sometimes long-term studies in order for the practitioner to make informed sometimes “on the spot” decisions in conjunction with players and coaches. This “working fast – working slow” concept is a relevant measurement issue because practitioners need to ensure that a given measurement of fatigue or performance can be interpreted quickly and accurately against the backdrop of random within-subjects variability, which may be quantified by the researcher in a well-designed reliability study. Any measurement of fatigue also needs to be interpreted with knowledge about how valid the measurement tool is to the ultimate outcomes of player performance, illness and/or injury. This knowledge is derived from well-designed and robust validity studies as well as prognostic-type studies in which fatigue is the predictor and injury rate, for example, is the outcome.

Measurement decisions should definitely not be based solely on whether a particular reliability or validity statistic, e.g. a correlation coefficient, is “statistically significant” or not. Similarly, published general qualitative thresholds of “excellent”, “good” or “moderate” measurement statistics may not always be fully informative for decision making purposes. For example, it has been thought in the past that a measurement tool can be deemed sufficiently reliable if the test-retest coefficient of variation (CV) is <10% and/or if a test-retest correlation is >0.8, but the rationale for these general thresholds has not been very clearly described in the past. The ideal process is for a given measurement of fatigue status to be interpreted
relative to both random within-subjects variability (thereby deriving the “minimal
detectable change”) as well as the minimum value that has been deemed to be
clinically or practically important; the “MCID”. It is important to consider that
the minimal detectable change might not be the same as the MCID. A minimum
detectable change is the smallest true change that has a reasonable probability of
being statistically significant. A properly-powered study should mean that the
minimum detectable effect is achieved, but this effect size might still not be clinically
or practically important.

Once the MCID has been selected and reliability/validity information is collected on a
relevant sample, this process is essentially probabilistic in nature, whereby one can estimate the chances that a given measurement exceeds the MCID. While useful spreadsheets exist to undertake this process the magnitude of the selected MCID is critical for the process of monitoring fatigue to be useful and informative in the field.

The quantification of an MCID is also imperative for accurate a priori sample size estimations in applied research. This sample size estimation is important for ethical and economic reasons. It may be unethical and wasteful of time and money for a researcher to collect data, possibly with invasive methods, if the sample size is too small to detect the MCID. A large sample size may also be unethical if a treatment effect could have been quantified with adequate precision using a much smaller sample.

An MCID might be relatively straightforward to select in “closed loop” sports like individual track cycling where the performance outcome to anchor to is clearly time, and objective indicators like the power produced by the rider are known to have a relatively precise association with this performance time. Changes in performance time (and associated indicators) may also be stable enough for the practitioner to link directly to meaningful changes in athlete ranking. Nevertheless, the selection of an MCID for team sport performance is more difficult due to the multi-component nature of these sports and the relatively high within-subject variability in physical performance between successive soccer-specific training drills and match-play. Furthermore, the use of highly demanding maximal performance tests, as possible closer surrogates of match performance, is difficult in team sports due to the limited recovery time between frequent competitions and matches. The magnitude of an MCID may also depend on whether it is considered from the perspective of the athlete, the coach or even the club’s financers/owners. For example, an MCID for a player may be the change in fatigue that leads to a perceived increase in match performance. The coach may be more interested in the MCID that confers the greatest probability of longer-term availability and the financers may be more interested in the MCID that confers both the former consequences as well as adding to the transfer value of the player. Obviously, all of these anchors are associated with each other, yet it is uncommon for these different perspectives to be considered together, using a Delphi method for example.

The approaches for selection of an MCID have mostly been considered in detail from a clinical trial perspective. Researchers and practitioners could gather information about the MCID using expert/end-user opinion, evidence synthesis and a pilot study, ideally by triangulating across these different approaches. Nevertheless,
although a certain difference derived from an evidence synthesis or pilot study might
be a realistic target, it might not necessary be clinically/practically important itself.

There are two main types of approaches for quantifying an MCID; anchor and
distributional (statistical). In the “anchor” approach, the measurement (or change in
measurement) is anchored to an associated change in another external measure of
change; the anchor variable. For example, a change in fatigue status could be
anchored to changes in illness, soft-tissue injury and/or match performances/rankings.
Ideally, the robustness of any relationship between an indicator variable (or exposure)
and the anchor should be quantified in prognostic-longitudinal type research. In
these types of studies, it is imperative to employ the most appropriate analysis
approaches, as highlighted recently by Finch and Marshall. These authors
recommended that epidemiological studies into sports injuries should use the
“Subsequent Injury Categorisation model” to make full use of all the longitudinal data
that are collected.

Distributional approaches for specifying an MCID can be based on the standard error
of measurement (typical error) and/or the between-subjects SD for the measured
variable (e.g. >0.2SDs). In the former approach, a measurement (or change in
measurement) is compared with an MCID that is larger than the random and
unavoidable within-subjects variability (standard error of mean). Additional decisions
in this approach are how much larger than random variability should the MCID be as
well as what level of statistical precision should be selected, although it has been
reported that a reasonable MCID approximates to 1SEM.

A related concept to the SEM is the “minimal detectable change”. For example, a
change in measurement could be deemed important if it exceeds twice the standard
deviation of differences (derived from a reliability study). The standard deviation of
differences can be estimated by multiplying the SEM by √2. This threshold of 2xSD
of differences is essentially the 95% limits of agreement described by Bland and
Altman.

Health economics may also be factored into the MCID selection process. This would
involve defining a threshold value for the cost of a change in performance or a
reduction in illness or injury that a coach or team owner is willing to pay. Then data
on the differences in costs, effects and harms can be all considered together to arrive
at an estimate of relative efficiency. It is common for changes in measurements to be
considered on a standardised scale, e.g. as a given fraction of the between-subjects
standard deviation. Various thresholds have been proposed for trivial, small, medium
and large effects. For injury applications, binary or survival (time-to-event)
outcome metrics (e.g. an odds, risk or hazard ratio) can be considered in a similar
way. Effect size thresholds for risk ratio values depend on the “nominal” event
proportion, i.e. that observed in the control group.

**Practical Applications & Future Direction**

Elite team sport athletes compete on a weekly basis and often 2-3 times per week in the
football codes. Whilst adhering to the fundamental measurement requirements outlined in
this review, prospective fatigue monitoring tools should also be non-invasive and time
efficient due to the large volume of players who may require assessing on any given day.
The tools should also minimize any additional loading due to the limited recovery time
available to team sport athletes during the competition phase. Recent surveys on current
trends of fatigue monitoring in high performance sport, highlight athlete self-report
questionnaires as the most frequently adopted tool, particularly customized designs
consisting of 4-12 items. The validity of these tools is also supported by a number of
reports highlighting their sensitivity to training and/or match load. However, the
efficacy of these tools is dependent upon a number of theoretical (inter-relations
between the measure, social environment and outcomes) and practical factors that need to
be addressed within the applied sports setting.

Research to date has mainly examined the sensitivity of prospective fatigue
monitoring tools to previous training and match load. Future work is required to
examine the degree to which a fatigue measure or change in fatigue measure promotes
subsequent changes in a relevant anchor such as performance, illness and injury. This
represents a move towards establishing the MCID for fatigue measures and in doing
so align with approaches adopted in clinical practice as outlined earlier in this
review. From a performance perspective, establishing pre-training or pre-match
MCID for fatigue measures would theoretically offer an indication on the quality of
the external output that might be produced. This would provide coaches with the
ability to make adjustments to the scheduled training or rotate the players from a
match perspective. However, application to match-play is unlikely to be feasible due
to the high inherent variability of match-play physical performance in team sports.

The use of more ‘closed loop’ sport-specific training drills where players may
possibly perform more consistent physical outputs may afford an opportunity to
establish pre-training MCID for fatigue measures.

There is now a growing body of literature highlighting links between increased athlete
loading (competition and training loads) and the incidence of soft-tissue injury and to
a lesser extent illness in team sports (see Drew et al., 2016 for a review of recent
work). Furthermore, internal loading may offer more accurate predictions of injury
risk than measures typically used to estimate the external load. In these studies,
internal load was largely estimated by multiplying total training or match session
duration with session ratings of perceived exertion. Future work, is required to
examine whether other measures of the internal load/fatigue status of the athlete may
further enhance the precision of these estimates. For example, a number of studies
have examined the relationship between psychological stress and injury in sports
though only few reports have examined these relationships in team sports using
athlete self-report assessments commonly used in practice. By examining the role
of various prospective measures of fatigue status we may enhance our ability to
observe changes in the athletes status, which may predispose them to illness and/or
injury. Although, fatigue experienced by team sport athletes is multifactorial in nature
and one single measure is insufficient in explaining athlete status. A combination of
subjective and objective measures is, therefore, more likely in order to quantify
fatigue status in elite team sport athletes.
Summary
Considering the increase in competition demands in elite team sports over recent years, the quantification of fatigue status has gained popularity amongst researchers and practitioners. Since maximal physical performance assessments (sprints, repeated sprints and maximal voluntary contractions) traditionally used to quantify recovery and fatigue in athletes are unsuitable in team sport environments due to their exhaustive and time-consuming nature, recent literature has demonstrated that quick, simple and non-invasive tools such as ASRM, ANS HR-indices, JROM and jump protocols which have been shown to be sensitive to changes in training load. Practitioners utilising such measures must consider the MCID when interpreting results to identify true sensitivity in athlete fatigue status and in turn, informed decisions alongside key stakeholders in elite team sport environments. This review has outlined the potential measures which may be used as a starting point by practitioners to monitor fatigue status in team sport athletes, however, future work is required to investigate the relationships between these measures of fatigue and global anchors such as performance, injury and illness.


680 60. Morgans R, Owen A, Doran D, Drust B, Morton JP. Pre-Match Salivary IgA in
681 Soccer Players From the 2014 World Cup Qualifying Campaign. *Int J Sports

684 fixture schedule induces a transient fall in salivary IgA in english premier

688 off-field “brains” provide a competitive advantage in professional football? *Br

690 63. Coutts AJ. Working Fast and Working Slow: The Benefits of Embedding

693 64. Atkinson G, Nevill AM. Statistical methods for assessing measurement error
694 (reliability) in variables relevant to sports medicine. *Sports Med*.
696 June 1, 2015.


700 Minimal clinically important differences in pharmacological trials. *Am J Respir

702 67. Rai SK, Yazdany J, Fortin PR, Aviña-Zubieta JA. Approaches for estimating
703 minimal clinically important differences in systemic lupus erythematosus.

706 69. Batterham AM, Atkinson G. How big does my sample need to be? A primer on

708 70. Kempton T, Sirotic A, Coutts A. Between match variation in professional

711 difference for a randomised controlled trial: DELTA (Difference ELicitation in

715 (PROGRESS) 1: a framework for researching clinical outcomes. *BMJ*.
716 2013;346:e5595.

717 73. Finch CF, Marshall SW. Let us stop throwing out the baby with the bathwater:

720 74. Wyrwich KW, Tierney WM, Wolinsky FD. Further evidence supporting an
721 SEM-based criterion for identifying meaningful intra-individual changes in

724 75. Bland JM, Altman DG. Measuring agreement in method comparison studies.