WANT TO SEE MY REPORT, COACH?

SPORT SCIENCE REPORTING IN THE REAL WORLD

– Written by Martin Buchheit, France

On the 9 March 2013, Sir Alex Ferguson delivered in the Irish Times probably one of the most encouraging ever message for sport scientists in football: "Sports science, without question, is the biggest and most important change in my lifetime. It has moved the game onto another level that maybe we never dreamt of all those years ago. Sports Science has brought a whole new dimension to the game". While such statements are gold for universities advertising sport sciences courses all over the world and for young students willing to embrace their carrier in elite clubs, the actual value of sport science may not always be rated as high in some elite clubs or federations[1]. Having an impact on the training programme, as a sport scientist, is anything but easy[1]. The way coaches and

athletes understand, accept and use sport science is highly variable and unpredictable. The path leading to effective sport science support is a is a long and winding road, with frequent stops and constant redirections required. Historically, many mistakes have been made while we learned about the veracity and usefulness of our data and the best ways to report and implement sports science in the elite sports setting. Among the different components of effective sport science support, the three most important steps are likely the following:

1. Having an appropriate understanding and analysis of the data; i.e. using the right metrics and statistics. The first consideration is the choice of the best variables, i.e. those can be trusted in terms of validity and relativity and that

can be useful to answer the questions that are actually asked by coaches and players. Second, working with relatively small numbers of athletes within a team setting as well as being unable to effectively control for many variables makes interpretation difficult with traditional analytical approaches such as Null Hypothesis Significance Testing (NHST, which includes 'p values' and 't-tests' for example). Over the last decade or so, however, great strides have been made in understanding and reporting the effects we have on our athletes and more valid and relevant approaches exist which are much easier to clinically interpret[2]. The modern practitioner working oblivious to these useful variables and analytical

TABLE 1

Limitations of NHST	Advantages of MBI
P values and in turn, study conclusions, are sample-size dependent (the greater the n, the lower the P), irrespective of the size of the effect. *The drop-out of a few athletes, or the lucky involvement of two more subjects can induce a 180° change in a study conclusion[5] (Table 2).*	***MBI can be applied to assess changes in individuals.*** *While individual score changes can be assessed in various ways (e.g. Z-scores[6], standard difference score[7]), MBI additionally allows us to assess the likelihood of these changes to be true for any given athlete, once the typical error of the test of interest and the SWC are known[8,9] (Figure 2).*
Significance doesn't inform on magnitude of effects, yet magnitude is what matters the most[10]. *With a large enough sample size, even very small, trivial or non-practical effects can appear significant (P<0.05). In practice, with 200 athletes showing a 0.01% improvement in performance, NHST would suggest that a nutritional supplement works, while the effects may in fact be negligible. Coaches and athletes are first interested in knowing what kind of performance benefits may be expected from the supplement (i.e. how much, the actual magnitude), and how likely this magnitude is to be of practical importance (i.e. likelihood of the effect to be greater than the SWC).*	***MBI allows authors to be honest with their sample size and better acknowledge trivial effects.*** *While a P>0.05 is often interpreted as a lack of an effect/difference, it is actually impossible to be confident that this is the right interpretation of the data analysis (sample size issue, type II error resulting from low statistical power). The beauty of MBI is that it allows us to distinguish between clear (confidence limits within the SWC) and unclear (confidence limits overlapping the SWC) trivial effects (Figure 1). This can't be achieved by NHST. An unclear effect/difference is not to be interpreted as lack of an effect, but suggests the need to increase sample size to improve precision.*
	MBI improves data visualisation. *MBI principles should be applied to graphical reports produced by sport scientists, where shaded trivial areas and confidence limits (or typical errors for individual data) are presented systematically to acknowledge the fact that not all changes are worthwhile and that some uncertainty always remains (Figures 1, 2 and 3).*

Table 1: Reasons why academics and practitioners should abandon null-hypothesis significance testing (NHST) and embrace magnitude-based inferences (MBI) (adapted from Buchheit, 2016[4]). SWC=smallest worthwhile change.

approaches could be considered incompetent, in my opinion, whereas a practitioner aware of these approaches but clinging to the past borders on disingenuous.

2. Offering attractive and informative reports via improved data presentation/ visualisation. Effectiveness in this step depends likely more on artistic skills and a creative mind than proper scientific knowledge and this is often overlooked in sport sciences programme. Day-to-day trials and errors are likely key in the search of the optimal data visualisation strategies.

3. Having appropriate communication skills and personal attitude to efficiently deliver these data and reports to coaches and athletes. This step is without doubt the most important of the process; there is however no training offered at universities for this. Nothing replaces experience, high personal standards and humility at this stage, which is generally developed over time.

The following sections will detail each of these three components.

COLLECTING AND UNDERSTANDING THE (RIGHT) DATA

The first important step to build a successful sport science system is to choose and work with the right data[3]. With the exponential rise in (micro) technology, collecting data from athletes has never been so easy. For every training session it is relatively easy to fully characterise both the external (e.g. tracking systems, encoders, force plates) and the internal load (e.g. heart rate, muscle oxygenation, sweat rate) placed on each athlete. However, technology per se might not be the solution; the foundations of successful sport science support are probably laid on the pitch first, when practitioners select the type of data that may help them to answer the questions that coaches and athletes have actually asked, in the way they collect these data, how they understand the limitations of each variable and how they analyse, report and utilise all

this information[3]. While validity/reliability studies are important in the search of the best variables, their practical usefulness should also not be overlooked, i.e. their ability to be used to impact on the training programme. This relates to 'interesting vs important' types of data. For example, measurement of maximal oxygen uptake vs maximal aerobic speed; only the latter can be used for training prescription.

Statistics are probably one of the most important aspects of sport science when it comes to using data to make decisions. Unfortunately, the statistical proficiency of most practitioners in the field is often insufficient to maximise the use of their data and in turn, impact meaningfully on training programmes. One of the main reasons for practitioners' lack of 'statistical efficiency' is that statistical lectures at university have, to date, exclusively sung the praises of NHST, which is:
• Not appropriate to answer the types of questions that arise from the field: as detailed in Table 1, the magnitude

	TABLE 2		
Type of data	*Example of data*	*Method to derive the SWC*	*Common SWC value*
Individual athlete performance	*Track and field events*	*1/3 of the performance coefficient of variation*	*~1% (0.1 s) for 100 m sprint time* *~3% (5 mins) for marathon*
Physical performance in team sports	*CMJ, sprint times, MAS*	*1) 1/5 of between-athlete SD* *2) performance clues e.g. based on empirical observations of direct performance benefits, such as a distance of 20-50 cm that one soccer player needs to be ahead of the opponent to win a ball*	*~2.5% (1 cm) for CMJ height* *~1.3% (0.2 km/h) % for MAS* *~1% (0.03 s) for 20 m sprint time*
Physiological data with no direct link to performance	*Heart rate variability*	*Factions/multiples of the within-athlete SD*	*Highly athlete-dependent* *The choice of the SD fraction/ multiple depends on the expected sensitivity (the greater the SWC, the more conservative the decisions)*
Physiological data with relationship with performance	*Submaximal HR*	*The actual change in this variable that relates to the smallest important change in performance*	*1% for submaximal HR*
Physical activity that has no direct impact on performance	*Distance covered during matches in team sports*	*Still debated* *1) 0.2 x between-athlete SD until new evidence is shown* *2) Interpretation of the magnitude left to the practitioners (Figure 3)*	*Likely depends on both tracking variables and intensity zones[15]*

Table 2: Suggested methods to derive the smallest worthwhile change[4]. For an exhaustive list of SWCs for different performance measures see the work of Hopkins[9] and Buchheit[16,17]. Change/differences of 1x, 3x, 6x and 10x SWC can be considered as small, moderate, large and very large, respectively[4]. SWC=smallest worthwhile change, CMJ=countermovement jump, MAS=maximal aerobic speed, SD=standard deviation.

of an effect is what matters the most to practitioners – P values don't inform this[4].

- Not appropriate to assess individuals, which is the core of elite athlete monitoring. In fact, conventional statistics allow analysis of population-based responses only (Table 1)[4].

As a valid alternative to NHST, clear analytical advances can be reached using magnitude-based inferences (MBI, Table 1). This 'new' statistical approach, driven largely by Will G. Hopkins' efforts over the past 15 years, has changed my life, both as an academic and practitioner in elite sport[11]. I personally hope that MBI is influential with other scientists, as it has been to me. While the debate will likely continue, MBI is today a well-established analytical approach in sports science and in other fields, particularly clinical medicine where practical/clinical significance often takes priority over statistical significance[4].

MBI is based on two simple concepts:
1. Changes/differences in any variable are systematically compared to a typical threshold representative of a smallest important or meaningful change (later to be termed the smallest worthwhile change, SWC[12]).
   a. **Why?** Not all changes are worthwhile/ meaningful. It is the magnitude of the change/difference that matters first: 'is the change larger/greater than the SWC? If yes, how many times greater?' In this context, change/differences of 1x, 3x, 6x and 10x SWC can be considered as small, moderate, large and very large, respectively[4].
   b. **How?** The most appropriate method to define it is however variable-dependent, which forces researchers to adopt a conscious process when analysing their data. "NHST is easy, but misleading. MBI is hard, but honest" (W.G. Hopkins, personal communication)[4].

Recommendations to calculate the SWC are provided in Table 2.

2. Instead of a classic 'yes or no' type response (NHST), the probabilities for these changes/differences to be 'real' (greater than the SWC) are reported.
   a. **More precisely:** chances are reported both quantitatively (e.g. 75/25/0 for percentage chances of greater/similar/ smaller magnitude than the SWC) and qualitatively (e.g. possibly, likely, very likely – Figures 1 and 2, and Table 3).
   b. **How?** These percentage chances and associated qualitative interpretations are generally set a priori (e.g. <1%, almost certainly not; 1 to 5%, very unlikely; 5 to 25%, probably not; 25 to 75%, possible; 75 to 95%, likely; 95 to 99, very likely; >99%, almost certain)
   c. **Practically:** these percentage chances can be obtained with only a few copy and paste manoeuvres using a specifically-designed spreadsheet freely available
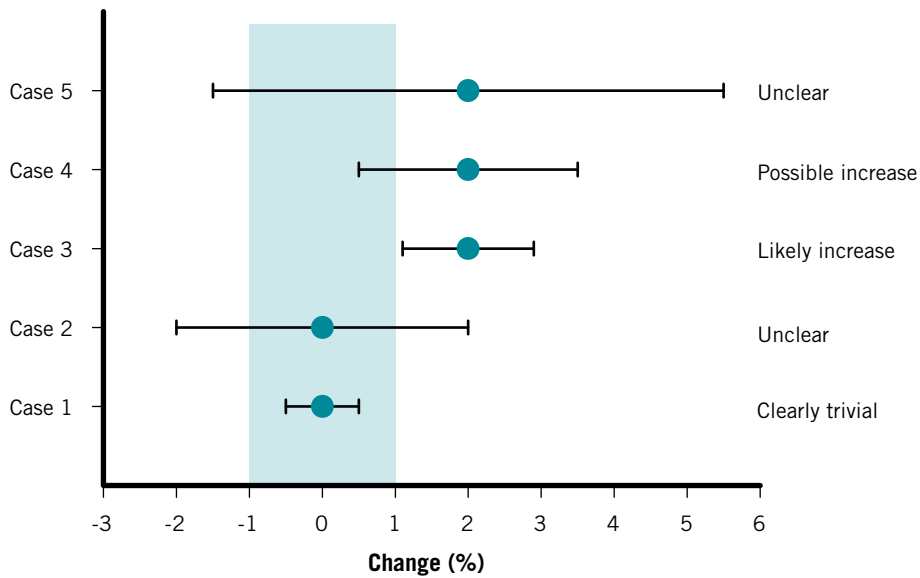
Figure 1: Example of possible decisions when interpreting changes using magnitude-based inferences. Note the clear vs unclear cases (based on confidence limits, in relation to the shaded trivial area), which is firstly, the beauty of magnitude-based inferences and, secondly, not possible via null hypothesis significance testing. Note also how, for clear effects, the likelihood of changes increases as the confidence limits shrink. Reprinted with permission from McCormack et al[5].
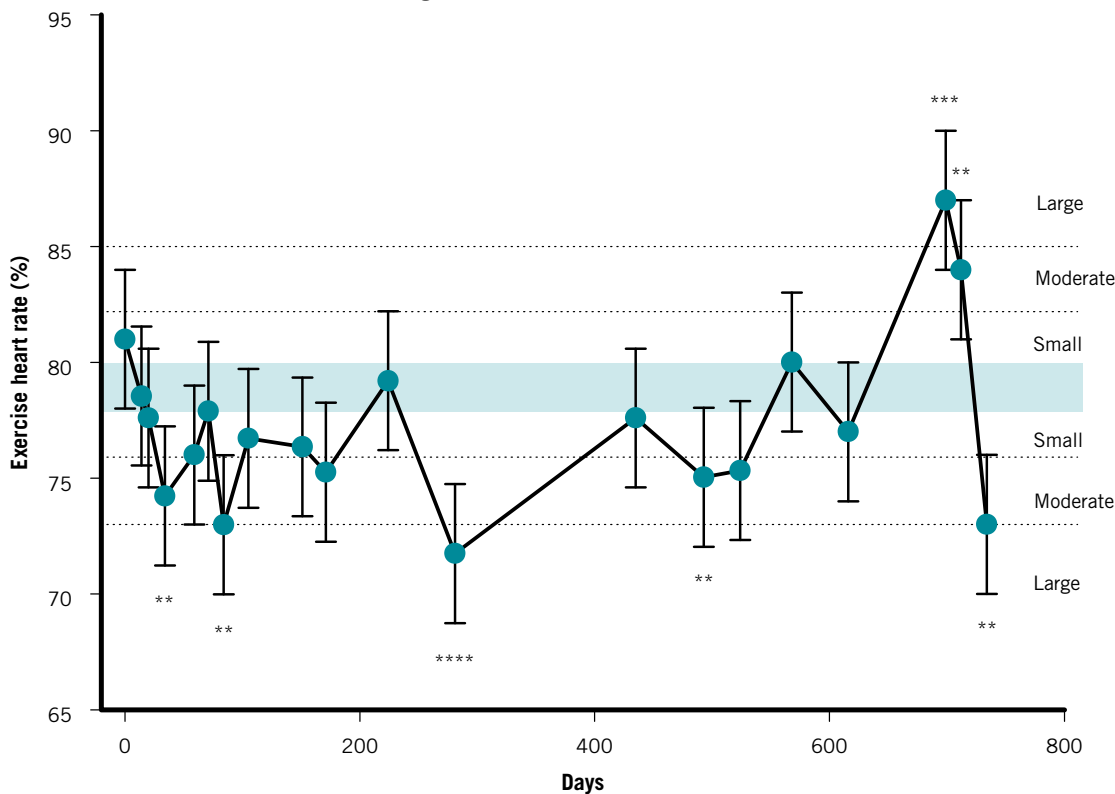

Figure 2: Individual changes in submaximal heart rate in a professional soccer player when running at 12 km/h throughout two competitive seasons (% of maximal heart rate). The shaded area represents trivial changes (1%)[3]. The error bars represent the typical error of measurement (3%)[3]. The number of * indicate the likelihood for the changes to be substantial, with ** referring to likely changes, and *** very likely changes. The magnitudes of the changes are set as multiples of the smallest worthwhile change (SWC); i.e. 1-3x (small), 3-6x (moderate) and large >6x SWC. Adapted from McCormack et al[5].

online[13,14]. Final decisions can then be translated into plain language when chatting with coaches: 'This attacker has very likely increased his sprinting speed. The magnitude of improvement should be enough for him to win a few more balls during matches.'

PRESENTING THE DATA

Similar to the aphorism that all roads lead to (and therefore from) Rome, the same data and results set can be presented in many ways (Figure 3). Once the relevant questions have been identified, the best variables have been selected and the appropriate statistics applied, the greatest challenge for sport scientists is to find the most efficient type of data visualisation and reporting to get their message across. Several considerations to optimise tables, graphs and content presentation are discussed below and illustrated in Table 3 and Figure 3.

1. Reports should be as simple and as informative as possible ('simple but powerful'):
   a. Limited to a few 'important' variables (those that can be used to answer the questions that coaches and athletes have actually asked and can have an impact on the programme).
   b. Extra decimals and 'noise' removed for clarity (Table 3).
   c. All text written horizontally for readability (Figure 3b).
   d. Labels added to graphs so that exact values can be seen too (graph for patterns, numbers for details, if required) (Figure 3b).
   e. Meaningful changes or differences highlighted to be seen at a glance (Figure 2) – with different possible levels of data analysis. Microsoft Excel's conditional formatting depicting MBI is a useful example (Table 2).

f. Including error bars where possible to acknowledge uncertainty (typical error of the measurement and confidence intervals for individual and average values, respectively) (Figure 2 and 3).

g. Using advanced visualisation tools such as Tableau or Microsoft BI. Although these require some training, they may be helpful to create aesthetically pleasing and advanced reports that may be more likely to catch coaches' and athletes' attention.

2. Format of the message should match with coach and athlete expectations, preferences and habits (which is linked to the search of the best delivery path, see below):

a. Visual vs verbal information.
b. Paper vs digital reports.
c. Quantitative vs qualitative interpretation.
d. Tables vs graphs (and types of graphs, e.g. bars vs radars etc.)

DELIVERING THE DATA

This last section is definitively less scientific than the previous two: it rather reflects personal views based on experiences and discussions with peers in the 'industry'. These ideas were recently summarised in an editorial for the *International Journal of Sports Physiology and Performance*[1]. While delivering the data is only one of the three steps highlighted in the present paper, it may be the most important. If sport scientists can't communicate with the coach, if they can't create interest and interactions with the coaches and players, then they obviously won't manage the get the message through (i.e. deliver) and their fancy reports with high quality stats will end up in the bin. The authors conclude: "Masters' degrees and PhD qualifications often are of little benefit in the quest of creating such a collaborative and productive environment. Understanding the specific codes of a sport or a very specific community of athletes takes many years. Having the respect and trust from high-profile athletes is often more a matter of personality and behaviour than scientific knowledge and skills. As described by the fantastic Dave Martin, we, sport scientists (monkeys) and coaches and athletes (felines and big cats) don't belong to the same species. We have

TABLE 3					
	Pre (cm)	Post (cm)		Pre (cm)	Post (cm)
Player 1	42	44	Player 1	42	44
Player 2	45	45	Player 2	45	45
Player 3	47	48	Player 3	47	48
Player 4	50	49	Player 4	50	49
Player 5	56	58	Player 5	56	58
Player 6	51	53	Player 6	51	53
Player 7	47	50	Player 7	47	50
Player 8	50	50	Player 8	50	50
Player 9	52	50	Player 9	52	50
Player 10	49	49	Player 10	49	49
Player 11	55	57	Player 11	55	57
Player 12	51	53	Player 12	51	53
Player 13			Player 13	54	56
Player 14			Player 14	46	46
Avg	49.6	50.5		49.6	50.6
SD	4.0	4.2		4.0	4.3
Standardised diff.	0.22 (small effect)			0.22 (small effect)	
P value - interpretation	0.06 – No effect			0.03 – Beneficial	
% for beneficial/ trivial/ harmful effect-interpretation	56/44/0 – Possibly beneficial			59/41/0 – Possibly beneficial	

Table 3: Effect of a nutritional supplement on jumping ability, which is used to illustrate the misleading nature of p values. In the present case, the inclusion of two more subjects (player 13 and 14), which doesn't even affect the group mean and standard deviation, induces a 180° change in the study conclusion using null hypothesis significance testing (not beneficial vs beneficial). In contrast, both the small magnitude of the effect (standardised changes >0.2[18], i.e. pre-post/pooled SD) and the overall data interpretation (inferences, % of chances for the supplement to have a beneficial effect) remain unchanged; they show the effectiveness of the nutritional supplement, irrespective of the sample size.
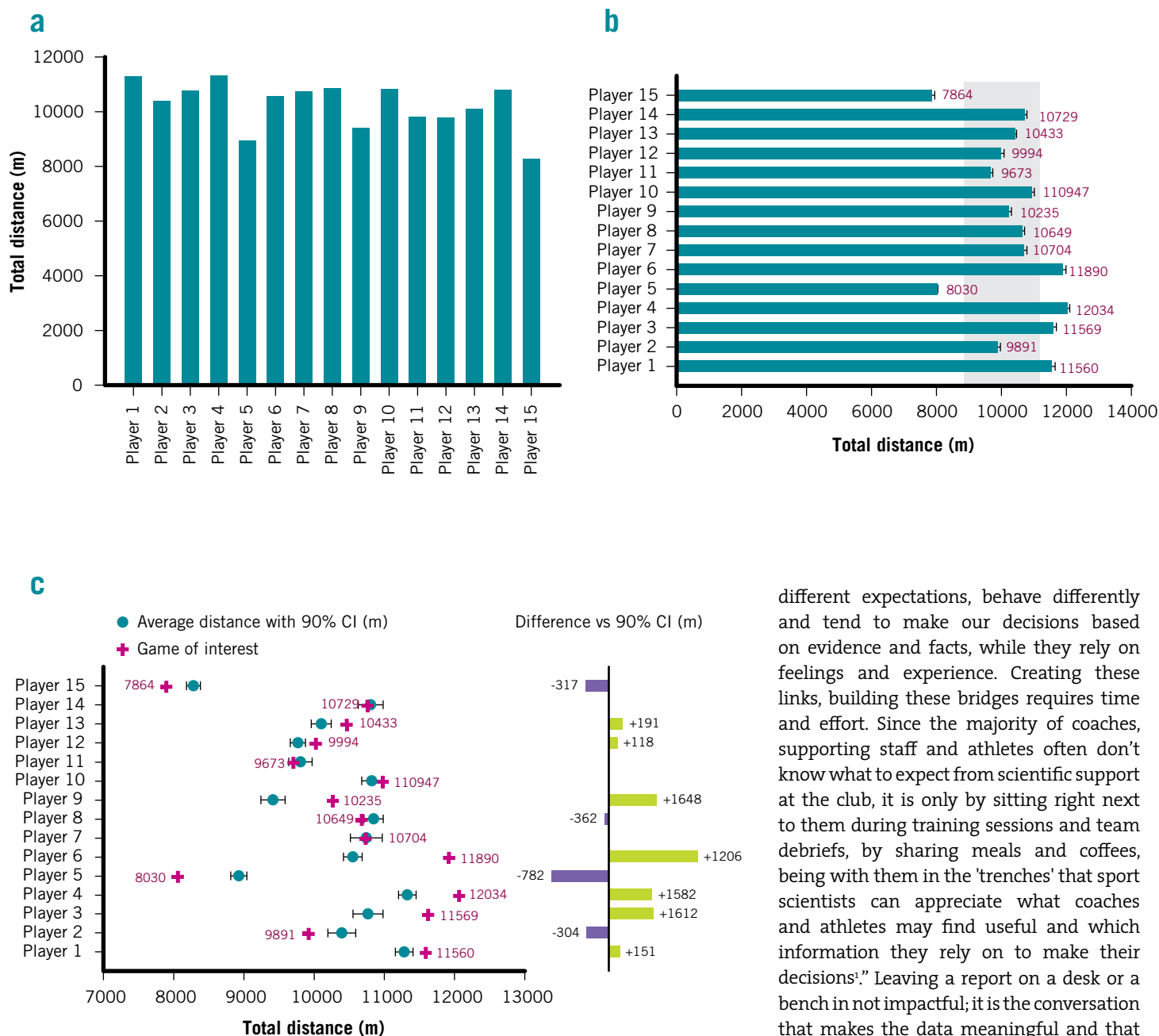
Figure 3: Illustration of various levels of data visualisation using distance covered during soccer matches as an example. Compared with (a), (b) is likely easier to read since all text is displayed horizontally and more informative: distance labels are provided on the side each the bars for more precision, error bars (typical error of the measurement, 1%) are added to reflect uncertainty of measurement and the shaded area represents team average ± standard deviation, which helps to visualise between-player differences. (c) highlights within-player differences for a given match of interest (red cross, players' top technical performance/impact on match result as rated by coaches) vs individual historical data (circle, with 90% CI). Since the most appropriate method to derive a smallest worthwhile change is still debated for such data (Table 2), the magnitude of the difference is provided in the actual unit (distance covered in meters that is outside the 90% confidence interval (90% CI, right part of the graph), and its interpretation is left to the practitioner. The take-home message from the graph is that there is no clear association between overall match outcome and total distance covered. CI=confidence intervals.

different expectations, behave differently and tend to make our decisions based on evidence and facts, while they rely on feelings and experience. Creating these links, building these bridges requires time and effort. Since the majority of coaches, supporting staff and athletes often don't know what to expect from scientific support at the club, it is only by sitting right next to them during training sessions and team debriefs, by sharing meals and coffees, being with them in the 'trenches' that sport scientists can appreciate what coaches and athletes may find useful and which information they rely on to make their decisions[1]." Leaving a report on a desk or a bench in not impactful; it is the conversation that makes the data meaningful and that can only occur once a relationship has been developed. Also, while having a strong character is often compulsory to survive in most places, open mindedness, humility and a form of kindness are probably some of the most important personality traits to develop to make an impact in this world. With these personal and social engagement skills in mind, it is not surprising that the majority of the most renowned researchers, sports scientists and performance managers to date have, in parallel to their academic journeys, exposed themselves deeply to the elite sport culture, either directly (as athletes) or indirectly (as coaches)[1]. Only those may have the ability to properly deliver data reports and influence decisions accordingly.

					TABLE 4		
Date	Raw HR	Rounded HR	Rounded HR + Excel conditional formatting	% HRmax + Excel conditional formatting	Changes + Excel conditional formatting (when >SWC)	Changes + Excel conditional formatting (when > SWC+TE)	Changes + Excel conditional formatting (based on inferences)
05/08/2014	151.470	151	151	81	2%	2%	+2%
19/08/2014	146.884	147	147	79	-1%	-1%	-1%
25/08/2014	145.104	145	145	78	-1%	-1%	-2%
08/09/2014	138.808	139	139	74	-5%	-5%	-5%**
03/10/2014	142.120	142	142	76	-3%	-3%	-3%
15/10/2014	145.652	146	146	78	-1%	-1%	-2%
28/10/2014	136.481	136	136	73	-6%	-6%	-7%**
18/11/2014	143.462	143	143	77	-2%	-2%	-3%
03/01/2015	142.777	143	143	76	-3%	-3%	-3%
23/01/2015	140.724	141	141	75	-4%	-4%	-4%
17/03/2015	148.116	148	148	79	0%	0%	+0%
13/05/2015	134.154	134	134	72	-7%	-7%	-8%***
14/10/2015	145.104	145	145	78	-1%	-1%	-2%
11/12/2015	140.314	140	140	75	-4%	-4%	-4%**
11/01/2016	140.861	141	141	75	-4%	-4%	-4%
24/02/2016	149.621	150	150	80	1%	1%	+1%
12/04/2016	143.990	144	144	77	-2%	-2%	-2%
04/07/2016	162.690	163	163	87	8%	8%	+8%***
17/07/2016	157.080	157	157	84	5%	5%	+4%**
08/08/2016	136.510	137	137	73	-6%	-6%	-6%**

Table 4: Example of various levels of data reporting using changes in submaximal heart rate responses to a standardised submaximal run. The level of clarity and usefulness increases from left to right. Individual changes in submaximal heart rate in a professional soccer player when running at 12 km/h throughout two competitive seasons (% of maximal heart rate). Adapted from Buchheit, 2016[4]. SWC=smallest worthwhile change (1%)[19], TE=typical error of measurement (3%)[19]. A change that is >SWC+TE has a 75% likelihood to be true[4]. The number of * indicates the likelihood for the changes to be substantial, with ** referring to likely changes, and *** to very likely changes, using a specifically designed spreadsheet freely available on the internet[12]. Data in the far right column are displayed in Figure 2.

CONCLUSION

The value and importance of sport science varies greatly between elite clubs and federations. Among the different components of effective sport science support, the three most important elements are likely the following:

1. Appropriate understanding and analysis of the data; i.e. using the most important and useful metrics only and using magnitude-based inferences as statistics. In fact, traditional null hypothesis significance testing (P values) is neither appropriate to answer the types of questions that arise from the field (i.e. assess magnitude of effects and examine small sample sizes) nor to assess changes in individual performances.

2. Attractive and informative reports via improved data presentation/visualisation ('simple but powerful').

3. Appropriate communication skills and personality traits that help to deliver data and reports to coaches and athletes. Developing such an individual profile requires time, effort and most importantly, humility.

> ## *Null hypothesis significance testing is easy, but misleading. Magnitude-based inferences are hard, but honest*

References

1. *Buchheit M. Chasing the 0.2. Int J Sports Physiol Perform 2016; 11:417-418.*

2. *Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. Int J Sports Physiol Perform 2006; 1:50-57.*

3. *Buchheit M, Simpson B. Player tracking technology: half-full or half-empty glass? . Int J Sports Physiol Perform 2016 [In press].*

4. *Buchheit M. The numbers will love you back in return – I promise. Int J Sports Physiol Perform 2016; 11:551-554.*

5. *McCormack J, Vandermeer B, Allan GM. How confidence intervals become confusion intervals. BMC Med Res Methodol 2013; 13:134.*

6. *McGuigan MR, Cormack SJ, Gill ND. Strength and power profiling of athletes: selecting tests and how to use the information for program design. Strength Cond J 2013; 35:7-14.*

7. *Pettitt RW. The standard difference score: a new statistic for evaluating strength and conditioning programs. J Strength Cond Res 2010; 24:287-291.*

8. *Al Haddad H, Simpson BM, Buchheit M. Monitoring changes in jump and sprint performance: best or average values? Int J Sports Physiol Perform 2015; 10:931-934.*

9. *Hopkins WG. How to interpret changes in an athletic performance test. Sportscience 2004; 8:1-7.*

10. *Cohen J. Things I have learned (so far). Am Psychol 1994; 45:1304-1312.*

11. *Buchheit M. Any Comments? 2013. Available from: www.herearemycomments.wordpress.com/. [Accessed 16 March 2016].*

12. *Hopkins WG. Statistical vs clinical or practical significance [Powerpoint presentation]. Sportscience 2002; 6. Available from: www.sportsci.org/jour/0201/Statistical_vs_clinical.ppt*

13. *Hopkins WG. Precision of the estimate of a subject's true value [Excel spreadsheet]. In: Internet Society for Sport Science. Sportscience 2000. Available from: www.sportsci.org/resource/stats/xprecisionsubject.xls2000 [Accessed November 2016].*

14. *Hopkins WG. A spreadsheet for deriving a confidence interval, mechanistic inference and clinical inference from a P value. Sportscience 2007; 11:16-20. Available from: www.sportsci.org/2007/wghinf.htm [Accessed November 2016].*

15. *Buchheit M, Allen A, Poon TK, Modonutti M, Gregson W, Di Salvo V. Integrating different tracking systems in football: multiple camera semi-automatic system, local position measurement and GPS technologies. J Sports Sci 2014; 32:1844-1857.*

16. *Buchheit M, Morgan W, Wallace J, Bode M, Poulos N. Physiological, psychometric, and performance effects of the Christmas break in Australian football. Int J Sports Physiol Perform 2015; 10:120-123.*

17. *Haugen T, Buchheit M. Sprint running performance monitoring: methodological and practical considerations. Sports Med 2016; 46:641-656.*

18. *Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. Med Sci Sports Exerc 2009; 41:3-13.*

19. *Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? Front Physiol 2014; 27:73.*

Martin Buchheit Ph.D.
Head of Performance
Paris Saint Germain Football Club
Paris, France
Contact: mbuchheit@psg.fr