SportRXiv

Part of the Society for Transparency, Openness and Replication in Kinesiology (STORK)

Preprint
not peer reviewed

Consensus meetings and statements are flawed by design: A narrative review with suggestions for improvements

Received: 23 April 2020
For correspondence: ian.shrier@mcgill.ca

Ian Shrier[1]

[1] Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, McGill University, 3755 Cote Sainte Catherine Road, Montreal, QC H3T 1E2

Please cite as: Shrier I. (2020). Consensus meetings and statements are flawed by design: A narrative review with suggestions for improvements. *SportRχiv*. https://doi.org/10.31236/osf.io/86t72

ABSTRACT

Consensus statements from the sport and exercise medicine community are now fairly common. More recently, the statements appear more prescriptive, strongly recommending particular approaches to

All authors have read and approved this version of the manuscript. This article was last modified on April 23, 2020

research or treatment. The most recent statement on methods for reporting sport injury surveillance studies included an extension to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines; STROBE guidelines are now official requirements for many journals. This suggests that investigators who use methods outside of these guidelines may have difficulty publishing their results.

The challenge is that by definition, consensus is not unanimity. Therefore, consensus recommendations are sometimes considered flawed at a later date. This is expected if we gain new knowledge. However, the consensus methods themselves may also inadvertently lead to a suppression of contrary but valid opinions.

The purpose of this narrative review it to propose a different model for consensus meetings and statements that embraces dissenting opinions, leading to increased transparency. In brief, the method is based on how Supreme Courts functions, allowing for both majority and one or more minority opinions. I illustrate how a consensus statement might be written using examples from four previous sport and exercise medicine consensus statements between 2005 and 2020. Such an approach will help ensure that clinicians, researchers and journals are not inappropriately influenced by recommendations from consensus statements, where experts continue to have important disagreements about the strength and interpretation of the evidence

INTRODUCTION

In February 2020, the International Olympic Consensus (IOC) published a consensus statement on methods for reporting sport injury surveillance studies.[1] The stated objective of the recent IOC statement is prescriptive: "to provide hands-on guidance to researchers on how to plan and conduct data collection and how to report data." [1] The consensus statement also includes an extension to the STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) reporting guidelines[2] called STROBE-SIIS; STROBE guidelines are now official requirements for many journals and one can reasonably assume journal reviewers and editors will apply STROBE-SIIS to future studies.

Researchers and journals need to be cautious before applying the recommendations of consensus statements in general. Despite best intentions, consensus statements have included recommendations that were later considered inappropriate. Although this is expected as we gain new knowledge, recommendations are sometimes contrary to knowledge available at the time the statement was adopted. One extreme example of inappropriate methods led to the call for a retraction[3] of the IOC consensus statement on managing load and injury in sport.[4]

Despite including leading researchers as part of the process, important deficiencies have occurred across many medical consensus statements. Therefore, I suggest this represents a fundamental and systemic problem with the underlying methods and reporting of the consensus meetings in general,

and not just within the sport and exercise medicine community. In brief, consensus statements imply consensus. At best, they represent decisions based on a majority of a committee. There are often differing opinions within the committee. However, similar to other consensus documents, the recent IOC consensus statement does not report dissenting opinions. Dissent and discussion are the foundation on how we interpret and then improve science; dissent needs to be embraced if we are to move forward appropriately.

As a solution, I propose that sport and exercise medicine adopt a previously published model that is similar to how a Supreme Court functions.[5] In this process, there is not one consensus. Rather, participants choose to align themselves with either a majority opinion, or with one or more minority opinions. When I have discussed this informally with colleagues, some have suggested the approach will cause confusion for readers. My perspective is that when there are disagreements within the research community, it is better for clinicians and researchers to be appropriately confused rather than inappropriately certain.

To illustrate the benefits of this approach, I first review the general methods used to develop consensus statements and highlight some challenges that lead to inaccuracies. In the subsequent sections, I consider some past consensus recommendations in the sport and exercise medicine field as the "majority opinion" and suggest what a "minority opinion" might look like within a final document using the Supreme Court model. The specific examples of inappropriate conclusions include how to categorize concussions,[6] restricting activity in women with female athlete triad,[7] managing load and injury in sport,[4] and methods for surveillance studies.[1]

General Methods for Consensus Statements

Although each consensus statement follows some unique processes, there are commonalities. The recent IOC statement on methods for injury surveillance[1] describes an 8-stage process. In many consensus statements, small working groups are formed of researchers who have published and maybe collaborated together to some extent in the area. The working groups write up texts among themselves. Once each text is written, it is sent to others (e.g. principal authors, a second sub-group) for editing / approval. The full text is written up and distributed to the entire group for approval, and then submitted for publication. Despite general acceptance of these processes, there are important areas where biases may be introduced.

First, consensus statements are written by those who were invited. In sport and exercise medicine, this generally includes clinicians and applied researchers, but methodologists and statisticians are less common. In addition, consensus meetings sometimes include young investigators or students of the organizers who do not have the same breadth of knowledge or experience, but would have equal votes to others in any "consensus of meeting experts". Finally, anyone organizing a consensus meeting already has their own perspective and agenda. Many people might avoid inviting those who they know have opposing views because it would simply make the meeting more difficult, and writing of the consensus document would take much longer. These three reasons may be why some

recommendations have been criticized for including seriously flawed analyses (see section on specific examples).

Second, most consensus meetings include a formal systematic review of the literature, but this has not always occurred in the sport and medicine exercise field. Where statements rely on research methods, epidemiological and statistical literature related to the issues being discussed must also be included. For example, the IOC Consensus statement on managing load and injury in sport was based on evidence synthesis of the acute:chronic workload ratio.[4] However, the analytical methods used in the original articles were *known to be flawed at the time of publication*,[8-10] and the evidence synthesis methods were also flawed.[11] This led to a series of articles criticizing the method,[11 12] and one call for retraction of the IOC statement itself.[3]

Third, most medical consensus statements do not define "consensus" criteria a priori.[13] In the recent IOC statement on harmonizing recording and reporting for injury, the methods state that "items were voted on to achieve a majority", which is >50%. I would suggest that 50.1% is not a strong enough endorsement to make prescriptive recommendations because "expert opinion" among the participants suggests the recommendation is just as likely to be incorrect as correct.

Fourth, discussions are held at consensus meetings and only later summarized and written up. There are rarely, if ever, official "votes" establishing how many participants agree with a text when comments or suggestions are subsequently made by co-authors. In many cases, authors are given explicit instructions that "approving" a text is not synonymous with "agreeing" to the text. Would you trust a recommendation more if it was approved by an 8:1 vote versus a 5:4 vote? Consensus statements need to transparent on how they determined what proportion of participants disagreed with particular parts of the text.

Concrete examples for a Supreme Court model

In each of the following examples, the majority opinion is a direct quote from a published sport and exercise consensus paper. The minority opinion represents some level of disagreement or elaboration on an important nuance or limitation within the majority opinion.

Concussions (2005)

The 2005 Concussion in Sport consensus statement suggested we categorize concussions as simple (resolves within 7-10 days) or complex (persistent symptoms).[6] This categorization was dropped unanimously a few years later at the 2008 consensus meeting.[14] However, even when the categorization was created in 2005, standard methods recommend that all information required for a categorization schema need to be available at the time the categorization is to be applied. When a concussion occurs, we do not know how long it will last. Therefore, we cannot diagnose a simple or complex concussion at the time of injury. In the following majority opinion, italics represents text that was not part of the direct quote but added for clarity.

Majority: "One of the key developments by the Prague Group is the understanding

that concussion may be categorized for management purposes as either simple (*resolves without complication over 7-10 days*) or complex (*loss of consciousness > 1 minute or prolonged recovery > 10 days*)."

Minority: Simple vs complex concussion categorization might be appealing to researchers who want to determine if risk factors (or treatment) for concussions leading to prolonged recovery are different from risk factors (or treatment) for concussions that resolve quickly. However, this analytical approach restricts data based on events that occur after the diagnosis, and this can lead to bias if the purpose is to determine causal risk factors.15 Further, it is of limited use for clinical management because one cannot generally apply the categorization at the time of injury or for the next 10 days.

Return to Play for Participants with Female Athlete Triad (2014)

Based on an earlier version[16] of the Strategic Assessment of Risk and Risk Tolerance (StARRT) model for return to play[17], this consensus statement[7] recommends using a cumulative risk assessment score (based on 6 individual factors including past history, symptoms and signs) to determine if a female athlete should be cleared for full activity. However, both the original[16] and StARRT[17] models propose that the magnitude of risk depends on the stresses applied during activity. If one followed the consensus recommendations, the magnitude of risk for an athlete who wants to play table tennis would be the same as an athlete who wants to run marathons.

Majority: "This cumulative risk stratification protocol is then translated into clearance and return-to-play guidelines for the Triad based on the athlete's cumulative risk score (figure 5). Future research is needed to assess if implementation of a risk stratification model results in improved outcomes for female athletes with Triad disorders."

Minority: The magnitude of risk is dependent on the activity being performed. Therefore, the "cumulative risk score" is a measure of bone health, not risk. In addition, the six individual factors are weighted equally in the score, whereas most clinicians would consider that an athlete with a bone mineral density Z-score between -1 and -2 is at much higher risk than an athlete with one previous stress fracture. We suggest clinicians consider the individual elements within this cumulative score as part of their overall decision-making process until there is empirical evidence supporting use of the score.

Managing load and injury in sport (2016)

The IOC consensus statement on managing load and injury in sport [4] recommended stratification of injury risk based on the acute:chronic workload ratio (ACWR). The ACWR was developed as a measure of change in activity: a recent (acute) change in activity compared to usual (chronic) activity. The statement reproduces a graph with a U-shaped curve suggesting that there is a "sweet spot" between ACWR of 0.8 and 1.3 that minimizes the risk of injury. Although the text of the document only discusses increases in acute load, the sweet spot is shown in the figure and explicitly mentioned in the associated infographic.[18] The implied interpretation is that decreasing activity by more than 20% (i.e. recent activity is 0.8 of usual activity) will result in the athlete being more at risk of injury compared to maintaining the same level of activity. There has never been a biological theory to support this statement, and the results are expected due to analytical methods that were known to be flawed at the time.[9-11 19]

Even if none of the authors were aware that the analytical methods were flawed (because there was no associated methodological literature review), the text for a majority and minority opinion in a Supreme Court model might have been (majority quote from Infographic article[18]):

> Majority: "Limit weekly increases of their training load to less than 10%, or maintain an acute:chronic load ratio within a range of 0.8 to 1.3, to stay in positive adaptation and thus reduce the risk of injuries"
>
> Minority: We agree with the majority opinion that injury risk increases as ACWR rises above 1.3. However, we cannot think of any biological reason why injury risk would acutely increase when activity is decreased (i.e. ACWR < 1), without a subsequent increase at a later time (i.e. ACWR >1). The results may have occurred by chance given the limited data available, or due to some unanticipated bias in the collection or analysis of data.

Methods for Recording and Reporting Injury Studies (2020)

The most recent IOC statement on reporting methods[1] includes several challenges. Its explicit purpose is to "harmonize" methods and analyses of sport medicine studies, and is being promoted as an extension to the STROBE statement. Therefore, it is likely to be required by some journals and the implications of un-recognized limitations is therefore much greater compared to the other consensus statements mentioned in this document. Therefore, this section discusses several challenges with the recommendations proposed so that future authors can publish their results using other methods if their study question and data require them.

Definitions

The first challenge is that several terms are not defined or operationalized for use. The recommended definition of injury in the consensus statement includes "transfer of kinetic energy" which requires motion. Although there is motion at the cellular and tissue level during an isometric contraction leading to damaged tissue, there is no motion at the joint level. The definition of injury also includes the word "damage". Is damage defined by the presence of bleeding or swelling, any rise in creatine kinase or only above a threshold, or something else? There is a body of medical literature discussing diseases versus illness, diagnoses versus incapacities, and so on that extend our traditional medical concepts of disease to include additional outcomes that are sometimes more meaningful to patients. Timpka et al. began to adapt these extended concepts to a sports framework in 2014 [20] and distinguished the following concepts: injury, trauma and incapacity along with more nuances for disease, illness and sickness. This work is consistent with the challenges that occur when we want to distinguish between recurrent injury and exacerbation,[21] and could also help with classifications when asymptomatic patients with known osteoarthritis or meniscal tears become symptomatic;[22] there are no new "injuries" from a medical perspective but these are events that we would normally want to keep track of and analyze properly in sport and exercise medicine. Further, although we traditionally define injury in sport and exercise medicine research as seeking all physical complaints, medical attention and time loss injuries, Bolling et al. found that some athletes consider an injury only if performance is affected, and others specifically mentioned that pain alone was not enough of a criteria to define an injury.[23] If we were to conduct a study to address research questions for these "clients" and were required to use the injury definition stated in the methods and reporting consensus statement[1] (or the updated OSTRC consensus statement[24]) as they suggest through the STROBE extension, then our results would clearly provide incorrect answers to the research question. When the objective of a consensus meeting is to be prescriptive, as this one was for the definition of injury, it is important for the relevant literature to be circulated, discussed and mentioned in the report.

There are challenges with other definitions provided as well. Non-contact is defined as "no contact from an external source", and "no evidence of disruption or perturbation of the player's movement pattern". Indirect contact is defined as an injury that results "from contact with other athletes or an object… The force is not applied directly to the injured area, but contributes to the causal chain leading to the health problem." However, this approach requires that one specify the most distal link in a chain of events (furthest away from the event) one is interested in. The example for indirect contact provided in the consensus statement is a skier who suffers a concussion "… after being knocked off balance hitting the gate with his knee." Now consider two skiers who both lose their balance because of ice, the head hits the snow hard and the athlete suffers a concussion. One athlete loses their balance because the ski slips and falls without hitting the gate. The other athlete has the same event and

although the skier maintains some control, they are still off balance resulting in their knee hitting the gate. According to these definitions, the first is a non-contact injury and the second is an indirect contact injury even though the initiating and final events in the chain (slip, head hitting snow) are the same.

Creating precise definitions is difficult and often requires many iterations and debate. The optimal definition will depend on whether the research is studying the most proximal cause of the pathology (e.g. head hitting the snow), or any one of the more distal causes (e.g. hitting gate, losing balance, poor sleep).

Time to recurrence

The IOC statement recommends that time to recurrence be recorded in days. An important analytical principle is that the denominator for any rate calculation should reflect the population-time at risk (known as risk set). Consider two athletes who are cleared to return to sport after an initial injury on Sunday, where one plays a sport only on Sundays, and the other plays a sport Monday, Wednesday and Friday. Both athletes get re-injured their first time playing. The athlete who plays once per week has a recurrence at 7 days, and the athlete who plays three times per week has a recurrence at 1 day. Is it fair to conclude the two sports have different recurrent injury rates when each athlete was reinjured on their first day back playing?

Majority: "Time to recurrence or exacerbation should be recorded in days (see 'Severity of health problems' section)."

Minority: The most appropriate measure for time to recurrence is necessarily dependent on the research question and the available data. We encourage authors to always report time related to "time at risk" (which might be days, or games depending on the sport), and to also report other common metrics when appropriate.

Multiple injuries at the same time

The consensus statement explicitly recommends injury prevalence and incidence should count multiple injuries occurring at the same time as one injury, and the severity should be considered the severity of the most severe injury. Therefore, if there were 10 events leading to simultaneous ankle and knee injuries over a season, the injury incidence would be 10 "injuries"/season overall, 10 ankle injuries/season, and 10 knee injuries/season for knee. Further, an ankle injury requiring only 1 week without activity will be considered severe if there is an associated fracture of the wrist.

Majority: "When one injury event results in more than one injury, the individual diagnoses should be recorded and classified separately. However, for injury incidence and prevalence reporting purposes this will be counted as one injury, and severity should be reported as the severity of the principal (most

severe) injury (see below for further explanation)."

Minority: Authors need to be clear if they are reporting incidence / prevalence / severity of events (e.g. falls) or incidence / prevalence / severity of injuries (e.g. knee sprain). If authors are reporting on injuries, then each pathology (ankle sprain, knee contusion) should count as one injury even if they occurred during the same event. If authors are reporting on events leading to injury, then the event leading to both an ankle and knee injury would count as only one event.

Fully recovered

The key distinction between exacerbations and subsequent injury is whether the initial injury had "healed/fully recovered." The consensus statement defines this as "fully available for training and competition", which is similar to a previous consensus statement.[25] However, the previous consensus statement[25] noted that athletes who continue to receive treatment after returning to full activity are still generally considered to be injured clinically. They proposed return-to-play criteria only as a pragmatic solution. I suggest the solution is dependent on the research question and we should not recommend one-solution-fits-all approach. As an example, the first question on the updated consensus statement regarding the Oslo Sport Trauma Research questionnaire[24] includes an answer choice "Full participation, but with (location) problems". Therefore, the two consensus statements are inconsistent in their recommendations of what might be considered an ongoing injury versus healed injury. This is understandable because the optimal definition for a study would depend on the data gathered (e.g. are symptoms being recorded), and the research question.

Majority: "Healed/fully recovered from injury (or illness) is defined as when the athlete is fully available for training and competition (see 'Severity of health problems' section)."

Minority: The optimal definition for healing is necessarily context dependent. Many athletes return to unrestricted activity but continue to receive treatment. In general, clinicians would not consider these injuries fully recovered. Further, patients with meniscal tears or osteoarthritis may be asymptomatic for months at a time but suffer from recurrent pain / swelling. These contexts illustrate the need for flexibility in reporting that is designed to optimally address the research question being asked, and to ensure we communicate efficiently with other interested stakeholders such as athletes and coaches who may interpret "healed" or "injury" differently than we have traditionally defined them.20 23 Options on how best to

record these “events” are beyond the scope of this discussion and are reported elsewhere.21 22

Summary

Consensus statements for treatment recommendations are usually based on systematic reviews with extensive readings and discussions. Sport and exercise medicine consensus statements that discuss methods need to be consistent with relevant epidemiological and statistical best practices. Failure to report dissenting opinions may lead to non-transparency and sub-optimal products, as illustrated by the concrete examples in this review. Such deficiencies are likely to hinder the advancement of sport medicine research, and by extension, sport medicine injury and illness prevention / treatment programs. Theoretically, the Supreme Court model may lead to more optimal results, but will still require having the correct mix of investigators’ experience and knowledge. This model has yet to be evaluated, and likely requires greater effort and time. Still, as the famous basketball coach John Wooden once said, “If you don’t have the time to do it right, when will you have the time to do it over?”

Funding information

This project was unfunded

REFERENCES

[1] R. Bahr, B. Clarsen, W. Derman et al. International Olympic Committee consensus statement: methods for recording and reporting of epidemiological data on injury and illness in sport 2020 (including STROBE Extension for Sport Injury and Illness Surveillance (STROBE-SIIS)). *Br J Sports Med* 2020;54(7):372-389.

[2] E. von Elm, D.G. Altman, M. Egger et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: Guidelines for Reporting Observational Studies. *PLoS Med* 2007;4(10):e296.

[3] F.M. Impellizzeri, S. Wookcock, A. McCall et al. The acute-chronic workload ratio-injury figure and its ‘sweet spot’ are flawed. Report dated: https://osf.io/gs8yu

[4] T. Soligard, M. Schwellnus, J.M. Alonso et al. How much is too much? (Part 1) International Olympic Committee consensus statement on load in sport and risk of injury. *Br J Sports Med* 2016;50(17):1030-41.

[5] A.D. SnidermanC.D. Furberg. Why guideline-making requires reform. *JAMA* 2009;301(4):429-31.

[6] P. McCrory, K. Johnston, W. Meeuwisse et al. Summary and agreement statement of the 2nd International Conference on Concussion in Sport, Prague 2004. *Clin J Sport Med* 2005;15(2):48-55.

[7] M.J. De Souza, A. Nattiv, E. Joy et al. 2014 Female Athlete Triad Coalition Consensus Statement on Treatment and Return to Play of the Female Athlete Triad: 1st International Conference held in San Francisco, California, May 2012 and 2nd International Conference held in Indianapolis, Indiana, May 2013. *Br J Sports Med* 2014;48(4):289.
[8] S. Williams, S. West, M.J. Cross et al. Better way to determine the acute:chronic workload ratio? *Br J Sports Med* 2016.
[9] G. AtkinsonA.M. Batterham. The use of ratios and percentage changes in sports medicine: time for a rethink? *Int J Sports Med* 2012;33(7):505-6.
[10] P. Menaspa. Are rolling averages a good way to assess training load for injury prevention? *Br J Sports Med* 2017;51(7):618-619.
[11] C. Wang, J.T. Vargas, T. Stokes et al. Analyzing activity and injury: Lessons learned from the acute:chronic workload ratio. *Sports Med* 2020 (In Press).
[12] L. Lolli, A.M. Batterham, R. Hawkins et al. Mathematical coupling causes spurious correlation within the conventional acute-to-chronic workload ratio calculations. *Br J Sports Med* 2019;53(15):921-922.
[13] I.R. Diamond, R.C. Grant, B.M. Feldman et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol* 2014;67(4):401-9.
[14] P. McCrory, W. Meeuwisse, K. Johnston et al. Consensus statement on Concussion in Sport 3rd International Conference on Concussion in Sport held in Zurich, November 2008. *Clin J Sport Med* 2009;19(3):185-200.
[15] M.A. Hernán, S. Hernandez-Diaz, J.M. Robins. A structural approach to selection bias. *Epidemiology* 2004;15(5):615-625.
[16] D.W. Creighton, I. Shrier, R. Shultz et al. Return-to-play in sport: a decision-based model. *Clin J Sport Med* 2010;20(5):379-385.
[17] I. Shrier. The strategic assessment of risk and risk tolerance (StARRT) framework for return to play decision making. *Br J Sports Med* 2015;49:1311-1315.
[18] T. Soligard, M. Schwellnus, J.M. Alonso. Infographic. International Olympic Committee consensus statement on load in sport and risk of injury: how much is too much? *Br J Sports Med* 2016;50(17):1042.
[19] L. Bornn, P. Ward, D. Norman. Training schedule confounds the relationship between acute: chronic workload ratio and injury. MIT Sloan Sports Analytics Conference; 2019; Boston.
[20] T. Timpka, J. Jacobsson, J. Bickenbach et al. What is a sports injury? *Sports Med* 2014;44(4):423-8.
[21] I. Shrier, R.J. Steele, M. Zhao et al. A multi-state framework for the analysis of subsequent injury in sport (M-FASIS). *Scand J Med Sci Sports* 2016;26:128-139.
[22] I. Shrier, B. Clarsen, E. Verhagen et al. Improving the accuracy of sports medicine surveillance – when is a subsequent event a new injury? *Br J Sports Med* 2017;51:26-28.
[23] C. Bolling, B.S. D., W. van Mechelen et al. How elite athletes, coaches, and physiotherapists perceive a sports injury. *Transl Sports Med* 2019;2:17-23.
[24] B. Clarsen, R. Bahr, G. Myklebust et al. Improved reporting of overuse injuries and health problems in sport: an update of the Oslo Sport Trauma Research Center questionnaires. *Br J Sports Med* 2020;54(7):390-396.

[25] C.W. Fuller, R. Bahr, R.W. Dick et al. A framework for recording recurrences, reinjuries, and exacerbations in injury surveillance. *Clin J Sport Med* 2007;17(3):197-200.