

Peer Review

Review of: "Determining the DNA of Sport Sciences Research Through Advanced Bibliometric Topic Modeling"

Stephane Perrey¹

1. University of Montpellier, Montpellier, France

This manuscript presents an ambitious and comprehensive bibliometric mapping of sport sciences research from 2001 to 2022 using advanced topic modeling derived from the Centre for Science and Technology Studies (CWT). The study's scope and emphasis on the multidisciplinary structure of the field constitute meaningful contributions. Overall, the paper is well-organized and grounded in methodological rigor, offering valuable empirical data for researchers and even policymakers. Below, I provide a balanced review that highlights strengths while offering concrete suggestions for improvement.

The article draws on a massive dataset (60,321 sport-science outputs; 4,159 global clusters), and the authors justify the need for a holistic overview (Background, p. 2–3). The argument that past bibliometric work on sport sciences has been fragmented is articulated clearly and motivates the study's objectives. This could be further emphasized in the introduction. The use of CWTS-based topic modeling is innovative within sport sciences. Most bibliometric work in this domain relies on co-word or co-citation analyses; the adoption of micro-level clusters is a methodological advancement (Methods, p. 4–5). The authors also provide accessible explanations of the clustering logic and micro-meso-macro levels. The inclusion of global and domain-specific maps (Figure 2; Figure 3) is particularly valuable. The results section offers well-structured findings, notably: growth in international collaboration (Table 1, Section 3.1), topic distribution across disciplines (Section 3.3), increasing concentration of publications over time (Figure 4A), detailed topical breakdowns (Tables 4, p. 16–21). Interestingly, these sections effectively demonstrate the field's evolution and the dominance of biomedical research. However, some points require clarification and further development. Please see the key points below.

While the study uses the WoS category as the operational boundary, more discussion is needed about the limitations of relying exclusively on WoS indexing (Section 4.1 touches on this but could be deeper). Authors are requested to add a subsection explaining: how WoS defines “sport sciences,” which subfields or journals may be systematically excluded, and how this may bias topic distribution (e.g., underrepresentation of coaching science, sociology of sport, or Indigenous sport studies). This contextualization will strengthen claims about interdisciplinarity.

The paper relies heavily on CWTS topic assignments, but readers would benefit from transparency about how overlapping topics were handled (Section 3.5 gives one example but is limited), and the potential misclassification of topics due to citation-based clustering. Can authors provide in Section 2 or 4 a short robustness assessment or sensitivity analysis, particularly for heavily interdisciplinary clusters? Even a brief comparison showing how many sport-science papers fall into clusters with <25% relevance would add clarity.

Figure 4A shows rising concentration in sport sciences topics between 2001 and 2018. The paper suggests plausible explanations (specialization, COVID effects), but these interpretations could be more analytically grounded. In Section 4.2, compare concentration trends between sport sciences and a baseline field (e.g., health sciences in general). Also, authors could consider breaking down Gini trends by macro-domain (biomedical vs. social sciences). This would reveal whether convergence is uniform or domain-specific. This addition would improve the strength of the discussion.

Regarding social science and humanities, section 3.3 notes that only 7.7% of publications fall into social sciences topics, but the discussion does not fully explain why this gap persists. In Section 4.3, authors could include specific recommendations for integrating SSH perspectives—such as encouraging journals in these areas to pursue WoS indexing or developing field-specific databases?

The tables in Section 3.5 are rich but also overwhelming. Some descriptions are overly general (“deep dive,” “holistic approach”), and the ordering of topics does not clearly align with the field’s structure. It might be relevant for major clusters (e.g., knee injuries, shoulder repair, concussion research) to include a short sentence stating why each is central to sport sciences (e.g., prevalence in athletes, economic relevance).

By grouping topics more explicitly by purpose (e.g., injury treatment, performance optimization, psychosocial factors, etc.), this will increase interpretability and reduce cognitive load for readers.

The dominance of biomedical topics (90.69% of publications; Table 3) is a key finding. However, the current discussion (Section 4.2) treats this primarily as a descriptive result. Can the authors add a short reflection on whether bibliometric indicators (citations, impact factors, indexing practices) systematically favor biomedical research?

Possibly, additional analyses that could strengthen the paper rely on longitudinal analysis of country specialization. Section 3.1 presents country-level outputs, but not how national topic portfolios have evolved. Adding this would enrich understanding of global dynamics.

Given that the sport sciences category expanded from 68 to 93 journals (Section 4.2), an analysis of whether new journals diversified or reinforced concentration would be insightful.

Finally, expanding the policy implications section with more concrete recommendations would also increase relevance for decision-makers.

Declarations

Potential competing interests: No potential competing interests to declare.