

Comparing statistical significance and practical importance in Iranian exercise physiology research: A meta-research study based on effect sizes and uncertainty

Daniel Tarmast ^{1*} - Ghasem Tarmast ²

1. Assistant Professor, Department of Sports Science, Faculty of Humanities, Pa.C., Islamic Azad University, Parand, Iran 2. Associate Professor, Department of Statistics, Faculty of Mathematical and Computer Sciences, Shahid Chamran University, Ahvaz, Khuzestan, Iran

(Received: 2026/01/26; Accepted: 2026/05/30)

Abstract

Distinguishing between statistical significance and practical importance remains a persistent challenge in interpreting findings in exercise physiology research. An overreliance on p-values may promote dichotomous interpretations, obscure statistical uncertainty, and potentially lead to suboptimal training or clinical decisions. The purpose of this meta-research study was to examine statistical reporting practices and to evaluate the correspondence between statistical significance and practical importance in Iranian exercise physiology articles. Using a systematic meta-research design, full-text articles published between 2019 and 2025 in Iranian exercise physiology journals were reviewed. After screening, 45 articles were included in the final analysis. Each article was coded for the reporting of p-values, exact p-values, effect sizes, uncertainty measures, and the availability of sufficient numerical data for effect size reconstruction. When possible, standardized mean differences corrected for small-sample bias were calculated along with their 95% confidence intervals, and results were categorized using a 2×2 framework (statistically significant vs. non-significant × practically important vs. not important). Most articles reported p-values, but rarely included effect sizes and confidence intervals. Estimation-based reanalyses showed a limited relationship between statistical significance and practical importance: some statistically significant results had trivial effects, while some non-significant results had potentially meaningful effects. These findings show the limits of relying on thresholds and the risk of confusing statistical significance with real impact. In conclusion, the results highlight the need to shift from p-value-centered reporting to an estimation-focused approach that emphasizes effect sizes and uncertainty. Such a shift is particularly important in exercise physiology, where applied decisions depend on an accurate understanding of both the magnitude and the precision of observed effects.

Keywords

Exercise physiology, statistical significance, effect size, practical importance, confidence interval.

*Corresponding Author: Email: dr.tarmast@iau.ac.ir

مقایسه معناداری آماری و اهمیت عملی در پژوهش‌های فیزیولوژی ورزشی ایران: یک مطالعه فراپژوهش بر پایه اندازه اثر و عدم قطعیت

دانیال تارمست*^۱ - قاسم تارمست^۲

۱. استادیار، گروه تربیت‌بدنی و علوم ورزشی، دانشکده علوم انسانی، واحد پرند، دانشگاه آزاد اسلامی، پرند، ایران

۲. دانشیار، گروه آمار، دانشکده علوم ریاضی و کامپیوتر، دانشگاه شهید چمران، اهواز، خوزستان، ایران

(تاریخ دریافت: ۱۴۰۴/۱۱/۰۶، تاریخ تصویب: ۱۴۰۵/۰۳/۰۸)

چکیده

تفاوت میان معناداری آماری و اهمیت عملی یکی از چالش‌های اساسی در تفسیر یافته‌های پژوهش‌های فیزیولوژی ورزشی است. اتکای صرف به مقدار احتمال می‌تواند به برداشت‌های ساده‌انگارانه از نتایج، نادیده گرفتن عدم قطعیت و در نهایت تصمیم‌گیری‌های تمرینی یا بالینی نامناسب منجر شود. هدف پژوهش حاضر، بررسی الگوی گزارش‌دهی مؤلفه‌های آماری و میزان هم‌راستایی معناداری آماری با اهمیت عملی در مقالات گرایش فیزیولوژی ورزشی منتشر شده در ایران بود. این مطالعه با طراحی فراپژوهش انجام شد و متن کامل مقالات علمی-پژوهشی گرایش فیزیولوژی ورزشی منتشر شده در بازه ۲۰۱۹ تا ۲۰۲۵ میلادی مورد بررسی نظام‌مند قرار گرفت. پس از غربال‌گری، ۴۵ مقاله وارد تحلیل نهایی شدند. برای هر مقاله، گزارش مقدار احتمال (مقدار p)، مقدار احتمال دقیق، اندازه اثر، شاخص‌های عدم قطعیت و کفایت داده‌های عددی برای بازسازی برآورد اثر کدگذاری گردید. در مواردی که داده‌ها اجازه می‌داد، اندازه اثر استاندارد شده تصحیح شده برای نمونه‌های کوچک و فاصله اطمینان ۹۵٪ محاسبه و نتایج در چارچوب ماتریس 2×2 (معنادار/غیرمعنادار $\times$ مهم/غیرمهم) طبقه‌بندی شد. یافته‌ها نشان داد گزارش مقدار احتمال در اغلب مقالات رایج است، اما گزارش اندازه اثر و فاصله اطمینان در درصد بسیار اندکی از مطالعات دیده می‌شود. نمونه‌های برآوردی نیز نشان دادند که معناداری آماری همواره با اهمیت عملی هم‌راستا نیست، به گونه‌ای که برخی نتایج معنادار از نظر عملی کوچک، و برخی نتایج غیرمعنادار از نظر عملی قابل توجه بودند. در مجموع، نتایج این فراپژوهش بر ضرورت گذار از گزارش‌دهی آستانه‌محور به رویکرد برآوردمحور، همراه با نمایش عدم قطعیت، در پژوهش‌های فیزیولوژی ورزشی تأکید می‌کند.

واژه‌های کلیدی:

اندازه اثر، اهمیت عملی، فیزیولوژی ورزشی، معناداری آماری، فاصله اطمینان.

مقدمه

است از نظر کاربردی متفاوت تفسیر شود، که می‌تواند به سطح آمادگی بدنی ورزشکار (۱۱، ۱۲)، رشته ورزشی (۱۳، ۱۴)، و اهداف (۱۵، ۱۶) در تحقیق مورد نظر بستگی داشته باشد. پژوهش‌های معاصر نیز بارها هشدار داده‌اند که اتکا به آستانه‌های قراردادی، مانند کمتر از ۰/۰۵ می‌تواند به روایت‌های اغراق‌آمیز یا حذف اثرات بالقوه مهم بی‌انجامد (۱۹-۱۷).

در گرایش فیزیولوژی ورزشی مشکل از آن‌جا آغاز می‌شود که مقدار p -در عمل، به یک کلید روشن/خاموش تبدیل می‌گردد: اگر معنادار شد، نتیجه مهم یا واقعی پنداشته می‌شود، و اگر غیرمعنادار شد، گویی اثر وجود ندارد (۷، ۲۰). این دوگانه‌سازی، چند خطای رایج می‌سازد: نخست، تفسیر صفر/یک که در آن مرزهای قراردادی جایگزین تفکر علمی می‌شوند، و دوم، خطای عدم معناداری=عدم اثر که در مطالعات با نمونه کوچک یا پراکندگی بالا بسیار فریبنده است (۷، ۲۱، ۲۲). سوم، زمینه را برای بروز رفتارهای انتخابی در تحلیل و گزارش‌نویسی فراهم می‌کند؛ از جمله دستکاری گزینشی انتخاب آزمون‌ها، متغیرها و مدل‌ها برای رسیدن به نتیجه دلخواه، پدیده‌ای که در ادبیات روش‌شناسی از آن با عنوان پی‌هکینگ و باغ‌راه‌های منشعب^۳ یاد می‌شود (۲۳). پژوهش‌های کلاسیک در روش‌شناسی نشان داده‌اند که انعطاف پنهان در انتخاب متغیرها، زمان‌های اندازه‌گیری، مدل‌ها و آزمون‌ها می‌تواند هر چیزی را معنادار جلوه دهد، حتی وقتی رابطه واقعی وجود ندارد (۲۳، ۲۴).

در پژوهش‌های علمی-ورزشی، تمایز میان معناداری آماری^۱ و اهمیت عملی نقشی تعیین‌کننده در درک واقعی ارزش یافته‌ها ایفا می‌نماید (۳-۱). یک تفاوت اندک احتمال دارد در متغیری فیزیولوژیک در سطح آماری معنادار جلوه کند، در حالی که همان تفاوت از منظر کاربردی و عملی برای بهبود عملکرد ورزشی ناچیز و بی‌اثر باشد (۴، ۵). در گرایش فیزیولوژی ورزشی، جایی که بدن انسان هم‌زمان آزمایشگاه و موضوع پژوهش است، خروجی تحلیل آماری نباید تنها یک برچسب باشد، زیرا تصمیمات تمرینی و سیاست‌گذاری درمانی و سلامت، از دل همین گزارشات بیرون می‌آیند (۶). در این میان، دو مفهوم که در بسیاری از مقالات به اشتباه هم‌معنا گرفته می‌شوند، معناداری آماری و اهمیت عملی-بالینی هستند. معناداری آماری به این می‌پردازد که داده‌های مشاهده شده تا چه اندازه با یک فرض صفر ازپیش‌تعریف شده ناسازگارند (۷).

با این حال، اهمیت عملی پرسش دیگری را دنبال می‌کند: اگر این تفاوت واقعی باشد، آیا برای ورزشکار، مربی، یا بیمار، آن قدر بزرگ هست که ارزش توجه و اقدام داشته باشد؟ این تمایز، به‌ویژه در ورزش و فعالیت‌های بدنی که گاهی یک تغییر کوچک می‌تواند مرز سکو و ناکامی را جابه‌جا کند، حیاتی است (۸). به‌عنوان نمونه، بهبود اندک در عملکرد هوایی (۹) یا کاهش بسیار کوچک در درصد چربی بدن (۱۰) ممکن

^۳ . The p-hacking and the garden of forking paths.

^۱ . Statistical significance.

^۲ . P-value.

در مواجهه با این کاستی روش‌شناختی در برخی پژوهش‌های فیزیولوژی ورزشی، ادبیات معاصر آمار توصیه می‌کند که به‌جای اتکاء به آستانه‌های قراردادی معنی‌داری، تفسیر یافته‌ها بر اندازه اثر و دامنه عدم قطعیت آن مبتنی باشد (۱۹، ۲۵). اندازه اثر نشان می‌دهد شدت تغییر یا میزان اختلاف تا چه حد بوده و فاصله اطمینان که در بیان دقیق‌تر می‌توان آن را بازه سازگاری نامید، روشن می‌کند این برآورد با چه میزان عدم قطعیت همراه است و کدام دامنه از مقادیر با داده‌های مشاهده‌شده سازگارتر به نظر می‌رسند (۱۷، ۲۱، ۲۶). چنین رویکردی، به‌جای صدور حکم‌های قطعی، به پژوهشگر اجازه می‌دهد قدرت اثر را هم‌زمان با مرزهای نادانسته‌ها گزارش نماید. این همان چیزی است که برای تصمیم‌گیری عملی در فیزیولوژی ورزشی، از طراحی تمرینات بدنی و برنامه‌های بازتوانی گرفته تا مداخلات درمانی، ضرورت دارد. در سال‌های اخیر، بیانیه‌ها و مقالات روش‌شناختی اثرگذار در رشته‌های مختلف علمی، بر ضرورت پرهیز از تقلیل تفسیر یافته‌ها به معناداری یا غیرمعناداری آماری تأکید کرده‌اند. در این رویکرد، گزارش اندازه اثر و شاخص‌های عدم قطعیت در کنار برآوردهای آماری، جایگزین تفسیرهایی می‌شود، که صرفاً بر مقدار p متکی هستند (۱۷، ۲۷).

در پژوهش‌های فیزیولوژی ورزشی نیز این توصیه‌ها صرفاً نظری نیستند و بسیاری از مطالعات معتبر در نشریات این حوزه، به‌طور فزاینده نتایج را همراه با اندازه اثر و فاصله اطمینان گزارش می‌کنند و حتی درباره میزان عدم قطعیت برآوردها، در صورت گسترده

بودن بازه‌های اطمینان، نیز بحث می‌کنند. برای مثال، در پژوهش کلر و همکاران (۲۰۲۴) در یک کارآزمایی تصادفی‌سازی‌شده، برآوردهای اثر تعامل زمان×گروه همراه با اندازه اثر استانداردشده و فاصله اطمینان ۹۵ درصد گزارش شده و آنان صراحتاً نقش عدم قطعیت را در تفسیر نتایج یادآور شده‌اند (۲۸). همچنین، مرورهای نظام‌مند و فرامطالعه‌های جدید درباره تمرین‌های تناوبی پرفشار نشان می‌دهند، گزارش میانگین تغییر به‌تنهایی کافی نیست و باید دامنه عدم قطعیت و ناهمگونی پاسخ‌ها نیز دیده و ارائه گردند (۲۹). افزون بر این، شواهدی وجود دارد که در حوزه پزشکی ورزشی و تمرین، سوگیری به نفع نتایج معنادار می‌تواند بر آنچه منتشر می‌شود سایه بیندازد و همین مسئله اهمیت نگاه فراپژوهشی به الگوهای گزارش‌دهی را بیشتر می‌کند (۳۰). با وجود این جریان جهانی (۷، ۱۵، ۱۷)، در ادبیات داخلی فیزیولوژی ورزشی ایران هنوز تصویر روشن و نظام‌مندی از کیفیت گزارش‌دهی مقدار p -دقیق، اندازه اثر و عدم قطعیت در دست نیست (۳۱-۳۳). شواهد گردآوری‌شده در چارچوب پیش‌ثبت و طراحی این فراپژوهش نشان می‌دهد تمرکز غالب مقالات بر ارائه مقدار p -بوده (۳۴) و گزارش منظم اندازه اثر و بازه‌های عدم قطعیت کمتر دیده می‌شوند (۳۵)، از اینرو، این وضعیت با توصیه‌های روش‌شناسی به‌روز هم‌راستا نیست (۷، ۸، ۱۵) و می‌تواند به برداشت‌های ساده‌انگارانه از یافته‌ها دامن بزند (۸).

نمونه‌هایی از متن مقالات نیز نشان می‌دهند چگونه زبان معنادار به‌سرعت به زبان اثرگذاری قطعی تبدیل می‌شود (۳۶-۳۸). برای مثال، در گزارشی که تفاوت‌ها

با $P < 0/05$ بیان شده، در نگارش علمی، هرگاه مقدار احتمال به صورت دقیق گزارش نشود، نمایش آن باید به شکل آستانه‌ای استاندارد ($p < 0/01$ یا $p < 0/05$) یا ($p < 0/001$) انجام گیرد، تا از القای دقت کاذب با صورت‌هایی مانند: $p = 0/0000$ پرهیز گردد. نتیجه‌گیری با عباراتی از جنس: بنابراین به نظر می‌رسد، به‌طور معناداری اثر گذاشته است، ادامه یافته، و مسیر استدلال بیش از آنکه بر اندازه تغییر و عدم قطعیت تکیه نماید، بر برچسب معناداری بنا شده است (۷، ۲۰، ۲۱). همین الگو، در حوزه‌های متنوعی از پژوهش‌های تمرینی و فیزیولوژیک می‌تواند سبب شود تغییرهای کوچک اما معنادار به عنوان دستاورد بزرگ تبلیغ شوند یا تغییرهای بالقوه مهم اما غیرمعنادار نادیده بمانند، به‌ویژه وقتی اندازه نمونه کم، پراکندگی زیاد، یا پیامدها متعدد باشند (۲۲، ۲۳، ۳۴). از این رو، این مطالعه فرای پژوهش با تمرکز بر مقالات فیزیولوژی ورزشی ایران به جای قضاوت‌های کلی، با سنجه‌های قابل بازرسی نشان دهد الگوی گزارش‌دهی و تفسیر نتایج چگونه شکل گرفته است (۳۹، ۴۰). پرسش محوری پژوهش آن است که مقالات، مقدار P را تا چه حد دقیق و کامل گزارش می‌کنند و اندازه‌اثر و فاصله اطمینان تا چه میزان حضور دارد یا دست‌کم از داده‌های مقاله قابل محاسبه است (۴۱). همچنین، نسبت الگوهای چهارگانه در یک چارچوب 2×2 چگونه توزیع می‌شود، یعنی در چه سهمی از نتایج، معناداری/غیرمعناداری با مهم/غیرمهم بودن از منظر عملی همسو یا ناهمسو است (۵، ۱۶، ۴۲). در نهایت، عدم قطعیت در چه سطحی گزارش می‌شود یا می‌توان آن را با اتکا به اطلاعات مقاله بازسازی کرد، تا فاصله

میان گزارش آماری و فهم تصمیم‌پذیر روشن‌تر گردد (۴۳). ضرورت انجام چنین کاری، صرفاً یک دغدغه روش‌شناسی انتزاعی نیست؛ بلکه تلاشی است برای بازگرداندن معنا به زبان عددها، تا آموزش دانشگاهی، داوری علمی و کاربست حرفه‌ای، بر پایه تفسیرهای سنجیده‌تر پیش رود (۳۹، ۴۴).

نتایج این مطالعه می‌تواند چارچوبی آموزشی برای پژوهشگران و دانشجویان فراهم آورد تا بدانند چگونه باید اندازه‌اثر و عدم قطعیت را گزارش و تفسیر نمایند. برای مربیان رشته‌های ورزشی، متخصصان علم‌تمرین، و تصمیم‌گیران حوزه سلامت و درمان، کمک کند میان اثر آماری و اثر قابل اتکا و قابل اقدام فرق بگذارند، تا توصیه‌های تمرینی و مداخله‌ای بر پایه شواهد معتبرتر استوار گردند (۲۱). همچنین برای داوران و سردبیران نشریات، معیارهایی عملی فراهم کند تا کیفیت گزارش‌دهی را منسجم‌تر ارزیابی کنند.

روش‌شناسی پژوهش

این پژوهش با طراحی فرای پژوهش انجام شد و به‌جای تولید داده اولیه، به صورت نظام‌مند بررسی می‌کند که پژوهش‌های گرایش فیزیولوژی ورزشی منتشرشده در ایران چگونه نتایج آماری خود را گزارش و تفسیر کرده‌اند و آیا مرزبندی معناداری آماری با اهمیت عملی/بالینی در گزارش‌ها رعایت شده است (۷، ۲۱). چارچوب مفهومی مطالعه همسو با پیام کلیدی ادبیات روش‌شناسی جدید است که هشدار می‌دهد تصمیم‌گیری و روایت علمی نباید به قاعده روشن/خاموش کمتر یا بیشتر از $0/05$ در نظر گرفته

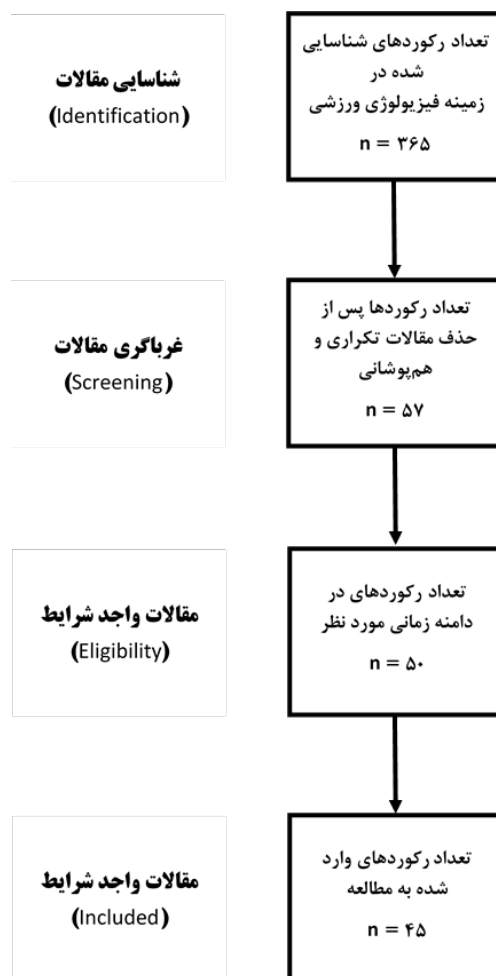
شود و به جای آن باید بر برآورد اثر و عدم قطعیت تکیه کرد (۷).

منبع داده، متن کامل مقالات گرایش فیزیولوژی ورزشی منتشرشده در نشریات علمی-پژوهشی علوم ورزشی ایران بود، که از تارنماهای رسمی نشریات دریافت و گردآوری شده بودند. واحد تحلیل مقاله تعریف شد و هر مقاله به عنوان یک مشاهده مستقل در نظر گرفته شد. در مواردی که یک پژوهش به صورت دو نسخه فارسی/انگلیسی در مجموعه وجود داشت، نسخه‌های تکراری تا حد امکان ادغام و به عنوان یک مقاله یکتا ثبت شدند. بر اساس روند غربال‌گری در همین مجموعه داده، تعداد ۳۶۵ رکورد اولیه شناسایی شد (نمودار ۱)، که پس از حذف همپوشانی‌ها به ۵۷ مقاله یکتا رسید، و نهایتاً ۴۵ مقاله در بازه زمانی سال‌های ۲۰۱۹ تا ۲۰۲۵ میلادی وارد تحلیل اصلی پژوهش حاضر گردیدند.

به منظور کاهش سوگیری ناشی از تعدد پیامدها و نیز نزدیک شدن به روایت اصلی هر مقاله، اگر مقاله چندین پیامد گزارش کرده بود، یک پیامد شاخص انتخاب شد (۴۵). معیار انتخاب پیامد شاخص مبتنی بر هدف، متغیر اصلی تأکیدی در عنوان/چکیده، و اولویت‌گذاری نویسندگان در بخش نتایج بود. بدین معنا که تلاش شد همان پیامدی انتخاب شود که نویسندگان، نتیجه‌گیری اصلی خود را بر پایه آن بنا کرده‌اند (۴۶). این تصمیم روش‌شناختی، با منطق آموزشی مطالعه نیز سازگار بود: وقتی حتی یک پیامد شاخص هم صرفاً با برچسب معناداری توصیف شود، زمینه برای تفسیر صفر/یک و نسبت دادن اهمیت

عملی به یک عدد فراهم می‌گردد، همان خطایی که در منابع گوناگون علمی به عنوان یکی از ریشه‌های اصلی برداشت‌های گمراه‌کننده معرفی شده بود (۷).

استخراج داده‌ها به صورت نظام‌مند انجام گرفت و برای هر مقاله، مؤلفه‌های گزارش‌دهی زیر به طور مستقل کدگذاری شد: (۱) وجود/عدم وجود گزارش مقدار p - (۲) گزارش مقدار دقیق p ، (۳) وجود و نوع اندازه‌اثر ارائه‌شده، (۴) گزارش شاخص‌های عدم قطعیت، و (۵) کفایت داده‌های عددی برای بازسازی اندازه‌اثر (۴۷-۴۹). در مواردی که گزارش نتایج آماری به صورت کلی یا آستانه‌ای ارائه شده بود، همان الگوی گزارش نیز به عنوان یک ویژگی گزارش‌دهی ثبت گردید (۵۰). هرگاه مقدار احتمال به جای عدد دقیق، به صورت آستانه‌ای گزارش شده بود، همان قالب آستانه‌ای ثبت شد (برای نمونه $p < 0.01$ یا $p < 0.05$ یا $p < 0.001$). همچنین، اگر در متن مقاله مقدار احتمال به اشتباه به صورت $p = 0.0000$ یا شکل‌های مشابه آمده بود، این مورد به عنوان گزارش آستانه‌ای تلقی شد (معادل $p < 0.001$) و در طبقه مقدار احتمال دقیق قرار نگرفت.



نمودار ۱. فلوچارت روند انتخاب مقالات واجد شرایط در گرایش فیزیولوژی ورزشی در بازه زمانی سال‌های ۲۰۱۹ تا ۲۰۲۵ میلادی

گزارش به صورت شفاف ثبت گردید، تا از یکی دانستن تغییر درون‌گروهی با اثر مداخله نسبت به کنترل پرهیز شود (۵۳، ۵۵). این رویکرد با منطق فراپژوهش حاضر همخوان است، زیرا هدف، توصیف واقعیت گزارش‌دهی است، نه بازنویسی مجدد نتایج به شکلی که در متن اصلی وجود نداشته است (۴۰، ۵۰).

برای سنجش اهمیت عملی، آستانه اصلی بر پایه اندازه‌اثر استاندارد شده تعریف شد، و معیار مهم عملی برابر با اندازه‌اثر $0.20 \leq$ در نظر گرفته شد (۵۶-۵۸). این آستانه به عنوان حداقل اثر کوچک اما بالقوه معنادار از نظر عملی به کار رفت، و همزمان یک تحلیل

نمونه‌هایی از این شیوه گزارش‌دهی در متن همین مجموعه مقالات مشاهده می‌شود. تعریف معناداری آماری در این مطالعه بر پایه آستانه رایج قرار گرفت ($P < 0.05$) (۷، ۵۱). با توجه به این که در برخی مقالات تنها آزمون‌های درون‌گروهی (پیش-پس) گزارش شده (۵۲)، و آزمون بین‌گروهی متناظر یا گزارش نشده بود، نوع آزمون (بین‌گروهی/ درون‌گروهی/مدل‌های تعاملی مانند گروه×زمان) ثبت شد (۵۳-۵۵). در تعیین معناداری آماری، در اولویت نخست از آزمون بین‌گروهی یا اثر تعامل استفاده شد، اگر فقط نتیجه درون‌گروهی موجود بود، همان وضعیت

حساسیت با آستانه سخت‌گیرانه‌تر | اندازه‌اثر | ۰,۰۵۵
انجام می‌شود، تا پایداری طبقه‌بندی‌ها نسبت به تعریف مهم بودن سنجیده شود (۵۶، ۵۷، ۵۹). این انتخاب، از یک سو با سنت رایج تفسیر اندازه‌اثرهای استاندارد شده همساز است، و از سوی دیگر با منطق کوچک‌ترین مقدار مهم که در ادبیات ورزش و سلامت به‌طور زمینه‌محور تعریف می‌شود، هم‌خوانی دارد (۶۰).
برای مثال، در برخی فراتحلیل‌های معتبر، حداقل بهبودهای بالینی برای شاخص‌هایی، مانند حداکثر مصرف اکسیژن به‌صورت صریح تعریف می‌شود (۶۱)، و برآوردها همراه فاصله اطمینان گزارش می‌گردد (۶۲)، تا مرز اثر مهم از اثر نامطمئن جدا شود (۶۴).

محاسبه اندازه اثر در سطح مقاله تنها زمانی انجام شد که برای پیامد شاخص داده‌های عددی کافی در دسترس بود (۶۵). برای مقایسه دو گروه در مرحله پس‌آزمون، از تفاوت میانگین استاندارد شده تصحیح شده برای نمونه‌های کوچک استفاده شد (۶۶).
در این روش، اختلاف میانگین‌های پس‌آزمون دو گروه بر انحراف معیار تجمیعی تقسیم گردید، و سپس ضریب تصحیح سوگیری ناشی از حجم نمونه کوچک اعمال شد تا برآورد به واقعیت نزدیک‌تر شود (۶۵، ۶۷).
برای سنجش عدم قطعیت پیرامون این برآورد، خطای معیار اندازه اثر محاسبه شد، و فاصله اطمینان ۹۵٪ بر پایه رابطه اندازه اثر $1,96 \times \text{خطای معیار}$ به دست آمد (۵۱، ۶۸).

این انتخاب عملیاتی بر دو ملاحظه استوار بود: نخست، استفاده از یک سنجه استاندارد شده موجب می‌شود نتایج مطالعات مختلف، حتی با مقیاس‌های

اندازه‌گیری متفاوت، در قالبی همسان و مقایسه‌پذیر گزارش شود (۶۷)؛ دوم، با تکیه بر محاسبه مستقیم از داده‌های عددی ارائه شده، وابستگی به گزارش‌های گزینشی نویسندگان کاهش می‌یابد (۶۸)، به‌ویژه در زمینه‌ای که گزارش منظم اندازه اثر و فاصله اطمینان در بخشی از متون فارسی (۶۹، ۷۰) هنوز فراگیر نشده است (۷۱، ۷۲).

هم‌زمان، برای تقویت جنبه آموزشی مطالعه و هم‌راستا کردن آن با استانداردهای بین‌المللی گزارش‌دهی، یادآوری می‌شود که در پژوهش‌های معاصر گرایش فیزیولوژی ورزشی (۱۹، ۳۱، ۳۲)، گزارش اندازه اثر همراه با فاصله اطمینان در کارآزمایی‌ها، مرورهای نظام‌مند و فراتحلیل‌ها رو به افزایش است (۵۰، ۷۱، ۷۳). در بسیاری از آثار جدید، برآورد اثر نه تنها با فاصله اطمینان ارائه می‌شود، بلکه گاهی آستانه‌های حداقل بهبود بالینی/عملی نیز تعریف می‌گردد (۶۱، ۶۲)، تا تفسیر نتایج از سطح معنادار/غیرمعنادار فراتر رود (۷۴). همچنین، در متون روش‌شناختی تصریح شده است که اتکای صرف به آزمون فرض در سطح ۵٪ برای تفسیر کافی نیست، و ارائه اندازه اثر در کنار نمایش عدم قطعیت (۷۵)، برای فهم دقیق‌تر بزرگی تغییر و حدود اطمینان‌پذیر آن ضرورت دارد (۴۲، ۵۰، ۶۸).

برای ساخت ماتریس چهارخانه 2×2 ، هر پیامد شاخص قابل طبقه‌بندی بر اساس دو محور گدگاری شد: محور نخست معنادار/غیرمعنادار آماری بر پایه $p < 0,05$ در نظر گرفته شده بود (۲، ۷)، و محور دوم مهم/غیرمهم عملی بر پایه آستانه اندازه اثر بود (۵۶). در

پاسخی است که ادبیات روش‌شناسی به بحران قانون‌های روشن/خاموش پیشنهاد می‌کند: شفافیت در گزارش، و تبدیل نتیجه از یک برچسب به یک برآورد قابل ارزیابی همراه با عدم قطعیت (۱۷، ۱۸، ۲۱، ۵۱).

یافته‌ها

در مسیر استخراج و پالایش مقالات، جریان انتخاب نمونه از یک مجموعه اولیه ۳۶۵ مقاله شناسایی شده آغاز شد، سپس پس از یکتاسازی و حذف موارد تکراری به ۵۷ مقاله یکتا و واجد شرایط رسید، و در نهایت ۴۵ مقاله برای تحلیل اصلی در بازه زمانی ۲۰۱۹ تا ۲۰۲۵ میلادی وارد مرحله تحلیل شد (نمودار ۱) (۷۶). این توالی، نقشه راهی است که نشان می‌دهد داده‌ها چگونه از انبوه متن‌ها به یک بدنه قابل اتکا برای ممیزی شیوه گزارش‌دهی و محاسبات اندازه‌اثر تبدیل شده‌اند.

در گام ممیزی کیفیت گزارش‌دهی، همان طوری که در جدول ۱ و نمودار ۲ گزارش شده‌اند، از میان مقالاتی که در این مرحله قابل ارزیابی بودند ($n=45$)، الگوی غالب چنین بود که مقدار p تقریباً همواره گزارش شده است (نمودار ۱) (۴۱، ۵۱). به‌طور مشخص، گزارش مقدار p در ۴۲ مقاله (۹۳/۳ درصد) مشاهده شد (نمودار ۲)، و گزارش مقدار دقیق p (یعنی ارائه یک مقدار عددی مشخص مانند $p = 0.12$ ، به‌جای گزارش آستانه‌ای مانند $p < 0.01$ یا $p < 0.05$ یا $p < 0.001$) در ۲۶ مقاله (۵۷/۸ درصد) ثبت گردید، و همچنین مواردی از نگارش‌های نادرست مانند

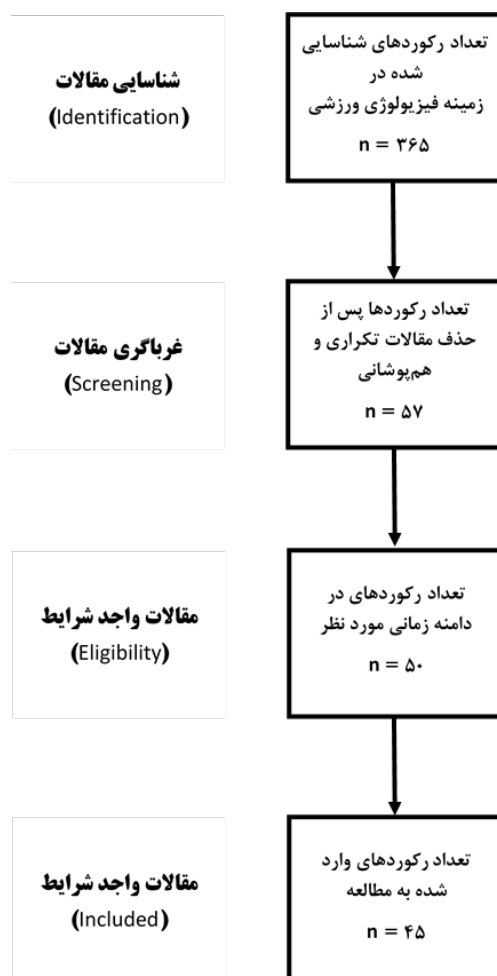
مواردی که اندازه‌اثر به‌دلیل نبود داده عددی کافی یا غیرقابل استخراج بودن جداول (برای مثال، جدول‌های تصویری یا گزارش‌های کیفی بدون عدد) قابل محاسبه نبود، وضعیت غیرقابل محاسبه ثبت شد (۶۸)، و آن موارد در تحلیل ماتریس 2×2 تنها در صورتی وارد می‌شوند، که طبقه‌بندی هر دو محور ممکن باشد (۶۸). با این حال، خود نرخ غیرقابل محاسبه بودن به‌عنوان یک شاخص مهم کیفیت گزارش‌دهی، در نتایج توصیفی گزارش می‌شود (۳۹، ۷۲، ۷۳)، زیرا همین ناتوانی در بازسازی عدم قطعیت، یکی از مصادیق اصلی فاصله گرفتن گزارش‌ها از رویکرد برآورد-محور است (۱، ۶۶).

برای افزایش قابلیت بازتولید و پاسخ‌گویی علمی، یک ردپای ممیزی دقیق برای هر عدد ثبت شد (۳۹، ۴۰). به این معنا که برای هر متغیر استخراجی یا هر عددی که در محاسبه اندازه‌اثر وارد می‌شد، صفحه و محل دقیق در پی‌دی‌اف (شماره صفحه و در صورت امکان شماره جدول/شکل) ثبت گردید (۳۹، ۴۰)، تا هر خواننده یا داوری بتواند مسیر رسیدن از متن مقاله به داده استخراج‌شده، و سپس به اندازه‌اثر و فاصله اطمینان را قدم‌به‌قدم پیگیری کند (۵۰). در راستای تقویت شفافیت روش‌شناختی و بازتولیدپذیری، داده‌های استخراج‌شده، محاسبات آماری و مستندات تکمیلی این پژوهش در سامانه چارچوب علم باز^۱ سپرده‌گذاری شده‌اند تا امکان دسترسی و بازبینی مستقل برای پژوهشگران فراهم باشد (<https://osf.io/wf8zr>). این رویکرد، دقیقاً همان

^۱. Open Science Framework (OSF)

هرگاه مقدار احتمال به‌جای عدد دقیق، به‌صورت آستانه‌ای گزارش شده بود، همان قالب آستانه‌ای ثبت شد (برای نمونه $p < 0.01$ یا $p < 0.05$ یا $p < 0.001$). همچنین، اگر در متن مقاله مقدار احتمال به‌اشتباه به‌صورت $p = 0.0000$ یا شکل‌های مشابه آمده بود، این مورد به‌عنوان گزارش آستانه‌ای تلقی شد (معادل $p < 0.001$) و در طبقه مقدار احتمال دقیق قرار نگرفت.

$p = 0.0000$ نیز، در صورت مشاهده در زمره گزارش‌های آستانه‌ای کُدگذاری شد (نمودار ۲).
استخراج داده‌ها به‌صورت نظام‌مند انجام گرفت و برای هر مقاله، مؤلفه‌های گزارش‌دهی زیر به‌طور مستقل کُدگذاری شد: (۱) وجود/عدم وجود گزارش مقدار- p ، (۲) گزارش مقدار دقیق p ، (۳) وجود و نوع اندازه‌اثر ارائه‌شده، (۴) گزارش شاخص‌های عدم قطعیت، و (۵) کفایت داده‌های عددی برای بازسازی اندازه‌اثر (۴۷-۴۹). در مواردی که گزارش نتایج آماری به‌صورت کلی یا آستانه‌ای ارائه شده بود، همان الگوی گزارش نیز به‌عنوان یک ویژگی گزارش‌دهی ثبت گردید (۵۰).



نمودار ۱. فلوچارت روند انتخاب مقالات واجد شرایط در گرایش فیزیولوژی ورزشی در بازه زمانی سال‌های ۲۰۱۹ تا ۲۰۲۵ میلادی

نمونه‌هایی از این شیوه گزارش‌دهی در متن همین مجموعه مقالات مشاهده می‌شود. تعریف معناداری آماری در این مطالعه بر پایه آستانه رایج قرار گرفت ($P < 0.05$) (۷، ۵۱). با توجه به این که در برخی مقالات تنها آزمون‌های درون‌گروهی (پیش-پس) گزارش شده (۵۲)، و آزمون بین‌گروهی متناظر یا گزارش نشده بود، نوع آزمون (بین‌گروهی/ درون‌گروهی/مدل‌های تعاملی مانند گروه×زمان) ثبت شد (۵۵-۵۳). در تعیین معناداری آماری، در اولویت نخست از آزمون بین‌گروهی یا اثر تعامل استفاده شد، اگر فقط نتیجه درون‌گروهی موجود بود، همان وضعیت گزارش به صورت شفاف ثبت گردید، تا از یکی دانستن تغییر درون‌گروهی با اثر مداخله نسبت به کنترل پرهیز شود (۵۳، ۵۵). این رویکرد با منطق فرایپژوهش حاضر همخوان است، زیرا هدف، توصیف واقعیت گزارش‌دهی است، نه بازنویسی مجدد نتایج به شکلی که در متن اصلی وجود نداشته است (۴۰، ۵۰).

برای سنجش اهمیت عملی، آستانه اصلی بر پایه اندازه‌اثر استاندارد شده تعریف شد، و معیار مهم عملی برابر با $|اثر| \leq 0.2$ در نظر گرفته شد (۵۸-۵۶). این آستانه به عنوان حداقل اثر کوچک اما بالقوه معنادار از نظر عملی به کار رفت، و همزمان یک تحلیل حساسیت با آستانه سخت‌گیرانه‌تر $|اثر| \leq 0.05$ انجام می‌شود، تا پایداری طبقه‌بندی‌ها نسبت به تعریف مهم بودن سنجیده شود (۵۶، ۵۷، ۵۹). این انتخاب، از یک سو با سنت رایج تفسیر اندازه‌اثرهای استاندارد شده همساز است، و از سوی دیگر با منطق کوچک‌ترین مقدار مهم که در ادبیات ورزش و سلامت

به‌طور زمینه‌محور تعریف می‌شود، هم‌خوانی دارد (۶۰). برای مثال، در برخی فراتحلیل‌های معتبر، حداقل بهبودهای بالینی برای شاخص‌هایی، مانند حداکثر مصرف اکسیژن به صورت صریح تعریف می‌شود (۶۱)، و برآوردها همراه فاصله اطمینان گزارش می‌گردد (۶۲، ۶۳)، تا مرز اثر مهم از اثر نامطمئن جدا شود (۶۴).

محاسبه اندازه اثر در سطح مقاله تنها زمانی انجام شد که برای پیامد شاخص داده‌های عددی کافی در دسترس بود (۶۵). برای مقایسه دو گروه در مرحله پس‌آزمون، از تفاوت میانگین استاندارد شده تصحیح شده برای نمونه‌های کوچک استفاده شد (۶۶). در این روش، اختلاف میانگین‌های پس‌آزمون دو گروه بر انحراف معیار تجمیعی تقسیم گردید، و سپس ضریب تصحیح سوگیری ناشی از حجم نمونه کوچک اعمال شد تا برآورد به واقعیت نزدیک‌تر شود (۶۵، ۶۷). برای سنجش عدم قطعیت پیرامون این برآورد، خطای معیار اندازه اثر محاسبه شد، و فاصله اطمینان ۹۵٪ بر پایه رابطه اندازه اثر $\pm 1.96 \times \text{خطای معیار}$ به دست آمد (۵۱، ۶۸).

این انتخاب عملیاتی بر دو ملاحظه استوار بود: نخست، استفاده از یک سنجه استاندارد شده موجب می‌شود نتایج مطالعات مختلف، حتی با مقیاس‌های اندازه‌گیری متفاوت، در قالبی همسان و مقایسه‌پذیر گزارش شود (۶۷)؛ دوم، با تکیه بر محاسبه مستقیم از داده‌های عددی ارائه شده، وابستگی به گزارش‌های گزینشی نویسندگان کاهش می‌یابد (۶۸)، به‌ویژه در زمینه‌ای که گزارش منظم اندازه اثر و فاصله اطمینان

در بخشی از متون فارسی (۶۹، ۷۰) هنوز فراگیر نشده است (۷۱، ۷۲).

هم‌زمان، برای تقویت جنبه آموزشی مطالعه و هم‌راستا کردن آن با استانداردهای بین‌المللی گزارش‌دهی، یادآوری می‌شود که در پژوهش‌های معاصر گرایش فیزیولوژی ورزشی (۱۹، ۳۱، ۳۲)، گزارش اندازه اثر همراه با فاصله اطمینان در کارآزمایی‌ها، مرورهای نظام‌مند و فراتحلیل‌ها رو به افزایش است (۵۰، ۷۱، ۷۳). در بسیاری از آثار جدید، برآورد اثر نه تنها با فاصله اطمینان ارائه می‌شود، بلکه گاهی آستانه‌های حداقل بهبود بالینی/عملی نیز تعریف می‌گردد (۶۱، ۶۲)، تا تفسیر نتایج از سطح معنادار/غیرمعنادار فراتر رود (۷۴). همچنین، در متون روش‌شناختی تصریح شده است که اتکای صرف به آزمون فرض در سطح ۵٪ برای تفسیر کافی نیست، و ارائه اندازه اثر در کنار نمایش عدم قطعیت (۷۵)، برای فهم دقیق‌تر بزرگی تغییر و حدود اطمینان‌پذیر آن ضرورت دارد (۴۲، ۵۰، ۶۸).

برای ساخت ماتریس چهارخانه ۲×۲، هر پیامد شاخص قابل طبقه‌بندی بر اساس دو محور کدگذاری شد: محور نخست معنادار/غیرمعنادار آماری بر پایه $p < 0.05$ در نظر گرفته شده بود (۲، ۷)، و محور دوم مهم/غیرمهم عملی بر پایه آستانه اندازه اثر بود (۵۶). در مواردی که اندازه اثر به دلیل نبود داده عددی کافی یا غیرقابل استخراج بودن جداول (برای مثال، جداول‌های تصویری یا گزارش‌های کیفی بدون عدد) قابل محاسبه نبود، وضعیت غیرقابل محاسبه ثبت شد (۶۸)، و آن

موارد در تحلیل ماتریس ۲×۲ تنها در صورتی وارد می‌شوند، که طبقه‌بندی هر دو محور ممکن باشد (۶۸). با این حال، خود نرخ غیرقابل محاسبه بودن به عنوان یک شاخص مهم کیفیت گزارش‌دهی، در نتایج توصیفی گزارش می‌شود (۳۹، ۷۲، ۷۳)، زیرا همین ناتوانی در بازسازی عدم قطعیت، یکی از مصادیق اصلی فاصله گرفتن گزارش‌ها از رویکرد برآورد-محور است (۱، ۶۶).

برای افزایش قابلیت بازتولید و پاسخ‌گویی علمی، یک ردپای ممیزی دقیق برای هر عدد ثبت شد (۳۹، ۴۰). به این معنا که برای هر متغیر استخراجی یا هر عددی که در محاسبه اندازه اثر وارد می‌شد، صفحه و محل دقیق در پی‌دی‌اف (شماره صفحه و در صورت امکان شماره جدول/شکل) ثبت گردید (۳۹، ۴۰)، تا هر خواننده یا داوری بتواند مسیر رسیدن از متن مقاله به داده استخراج‌شده، و سپس به اندازه اثر و فاصله اطمینان را قدم‌به‌قدم پیگیری کند (۵۰). در راستای تقویت شفافیت روش‌شناختی و بازتولیدپذیری داده‌های استخراج‌شده، محاسبات آماری و مستندات تکمیلی این پژوهش در سامانه چارچوب علم باز^۱ سپرده‌گذاری شده‌اند تا امکان دسترسی و بازبینی مستقل برای پژوهشگران فراهم باشد (<https://osf.io/wf8zr>). این رویکرد، دقیقاً همان پاسخی است که ادبیات روش‌شناسی به بحران قانون‌های روشن/خاموش پیشنهاد می‌کند: شفافیت در گزارش، و تبدیل نتیجه از یک برچسب به یک برآورد قابل ارزیابی همراه با عدم قطعیت (۱۷، ۱۸، ۲۱، ۵۱).

^۱. Open Science Framework (OSF)

یافته‌ها

در مسیر استخراج و پالایش مقالات، جریان انتخاب نمونه از یک مجموعه اولیه ۳۶۵ مقاله شناسایی شده آغاز شد، سپس پس از یکتاسازی و حذف موارد تکراری به ۵۷ مقاله یکتا و واجد شرایط رسید، و در نهایت ۴۵ مقاله برای تحلیل اصلی در بازه زمانی ۲۰۱۹ تا ۲۰۲۵ میلادی وارد مرحله تحلیل شد (نمودار ۱) (۷۶). این توالی، نقشه راهی است که نشان می‌دهد داده‌ها چگونه از انبوه متن‌ها به یک بدنه قابل اتکا برای ممیزی شیوه گزارش‌دهی و محاسبات اندازه‌اثر تبدیل شده‌اند.

در گام ممیزی کیفیت گزارش‌دهی، همان طوری که در جدول ۱ و نمودار ۲ گزارش شده‌اند، از میان

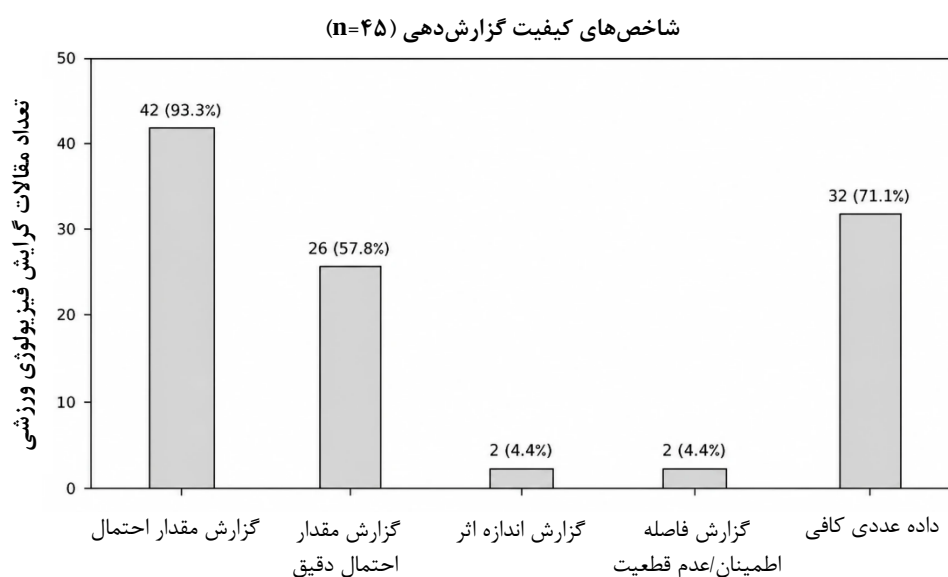
مقالاتی که در این مرحله قابل ارزیابی بودند ($n=45$)، الگوی غالب چنین بود که مقدار p تقریباً همواره گزارش شده است (نمودار ۱) (۴۱، ۵۱). به‌طور مشخص، گزارش مقدار p در ۴۲ مقاله (۹۳،۳ درصد) مشاهده شد (نمودار ۲)، و گزارش مقدار دقیق p (یعنی ارائه یک مقدار عددی مشخص مانند $p = 0.012$)، به‌جای گزارش آستانه‌ای مانند $p < 0.01$ یا $p < 0.05$ یا $p < 0.001$ در ۲۶ مقاله (۵۷،۸ درصد) ثبت گردید، و همچنین مواردی از نگارش‌های نادرست مانند $p = 0.000$ نیز، در صورت مشاهده در زمره گزارش‌های آستانه‌ای کُذگذاری شد (نمودار ۲).

جدول ۱. شاخص‌های کیفیت گزارش‌دهی در مقالات قابل ممیزی ($n=45$)

شاخص گزارش‌دهی	تعداد (n)	درصد (%)
گزارش مقدار احتمال	۴۲	۹۳/۳
گزارش مقدار احتمال دقیق	۲۶	۵۷/۸
گزارش اندازه‌اثر	۲	۴/۴
گزارش فاصله اطمینان/شاخص‌های عدم قطعیت	۲	۴/۴
وجود داده عددی کافی برای بازسازی اندازه‌اثر	۳۲	۷۱/۱

در مقابل، دو مؤلفه‌ای که برای فهم شدت اثر و عدم قطعیت حیاتی‌اند، یعنی اندازه‌اثر و بازه اطمینان/شاخص‌های عدم قطعیت، به‌طرز معناداری کم‌فروغ بودند (جدول ۱ و نمودار ۲) (۱۹، ۵۱). تنها ۲ مقاله (۴،۴ درصد) به اندازه‌اثر اشاره/گزارش داشتند و فقط ۲ مقاله (۴،۴ درصد) بازه اطمینان یا گزارشی مستقیم از عدم قطعیت ارائه کرده بودند (نمودار ۲). با این حال، وجود داده عددی (مانند میانگین $\pm$ انحراف

معیار یا صورت‌های مشابه) در ۳۲ مقاله (۷۱،۱ درصد) دیده شد (۴۳، ۵۶، ۶۸). نکته‌ای که نشان می‌دهد مسئله، صرف نبود داده نیست، بلکه غالباً چگونگی ارائه داده و ترجمه آن به برآورد اثر و عدم قطعیت است (۳۹، ۵۱).



نمودار ۲. شاخص‌های گزارش‌دهی در مقالات گرایش فیزیولوژی ورزشی

این مرحله تنها آن دسته از مقایسه‌ها وارد تحلیل شدند، که هم پیامد شاخص در متن مقاله به‌طور روشن قابل تعیین بود (۴۶). همچنین، داده‌های پس‌آزمون دو گروه به اندازه‌ای کفایت داشت، که بتوان تفاوت میانگین استاندارد شده تصحیح شده برای نمونه‌های کوچک را محاسبه نمود، و برای آن فاصله اطمینان ۹۵٪ را بر پایه خطای معیار به دست آورد (۴۳، ۶۸). بر اساس مواردی که در این فاز قابل محاسبه و تأیید بودند، دو نمونه روشن حاصل شد، که هر دو از منظر اهمیت عملی، بر مبنای آستانه از پیش تعیین شده $|t| \leq 2.0$ ، در محدوده مهم قرار گرفتند (۳، ۵، ۲۰). با این حال، از منظر معناداری آماری وضعیت یکسانی نداشتند. همین ناهمخوانی ظریف میان برجسب آماری و برآورد اثر، نقطه‌ای است که در آن تمایز میان معناداری آماری و اهمیت عملی

این تصویر در پیمایش مرحله‌ای فایل تجمیعی نیز بازتاب داشت. در متن مقالات منتخب، گزارش مقدار- p بر گزارش اندازه‌اثر و بازه اطمینان پیشی گرفته است (۳۴، ۴۴، ۷۳). به بیان دیگر، در حالی که در هر مرحله چندین مقاله دارای آزمون معناداری بودند، گزارش بازه اطمینان تنها در تعداد بسیار محدودی از مقالات و به‌صورت پراکنده ظاهر شد (۳۴، ۷۳). چنین الگوی مرحله‌به‌مرحله‌ای، همان منطق ردیابی ممیزی را تقویت می‌کند (۳۴، ۳۹): برای هر عدد کلیدی، محل دقیق آن در متن/جدول ثبت می‌شود، تا هر خواننده یا داور بتواند مسیر استخراج را بازسازی کند (۳۹، ۵۰، ۶۸)، و اختلاف میان گزارش و قابلیت محاسبه به‌صورت شفاف قابل بررسی و پیگیری باشد (۳۹، ۴۰).

در بخش یافته‌های برآورد محور و طبقه‌بندی ماتریس 2×2 (معنادار/غیرمعنادار $\times$ مهم/غیرمهم)، در

نه در سطح شعار، بلکه در سطح تصمیم‌گیری واقعی آشکار می‌شود (۲، ۷، ۲۱).

در جدول ۲ به نمونه‌های محاسبه‌پذیر و تأییدشده برای طبقه‌بندی ۲×۲ اشاره شده است (۴۳، ۶۵). در نمونه نخست، پیامد شاخص $\dot{V}O_{2max}$ در مقایسه مکمل ترکیبی در برابر کنترل با اندازه‌اثر ۰/۲۹۷ به‌دست آمد و مقدار p -گزارش‌شده ۰/۰۱۲ بود (جدول ۲). این نمونه، از منظر طبقه‌بندی آموزشی پژوهش حاضر، نماینده وضعیتی است که در آن اثر از آستانه

عملی عبور می‌کند، و هم‌زمان گزارش مقاله نیز به معناداری آماری اشاره دارد. یعنی همان ربع معنادار و مهم. با این حال، نکته ممیزی این بود که در برخی مقالات، مقدار p -الزاماً به مقایسه بین‌گروهی پس‌آزمون مربوط نیست و ممکن است به اثر زمان، درون‌گروهی، یا آزمون‌های دیگر گره خورده باشد (۵۳، ۵۵). بنابراین، ثبت نوع آزمون در کنار p ، برای جلوگیری از برداشت‌های شتاب‌زده، یک جزء کلیدی از گزارش‌دهی قابل دفاع محسوب می‌شود (۴۱، ۵۰، ۵۱).

جدول ۲. نمونه‌های محاسبه‌پذیر و تأییدشده برای طبقه‌بندی ۲×۲ (معناداری آماری×اهمیت عملی)

پیامد شاخص	مقایسه	اندازه اثر	مقدار احتمال	معناداری آماری	اهمیت عملی	طبقه‌بندی ۲×۲
حداکثر اکسیژن مصرفی ($\dot{V}O_{2max}$)	مکمل ترکیبی در برابر کنترل	۰/۲۹۷	۰/۰۱۲	معنادار	مهم	معنادار×مهم
CTRP-۱۲	تمرین مقاومتی در برابر کنترل	۰/۶۲۱	۰/۲۵۲	غیرمعنادار	مهم	غیرمعنادار×مهم

در نمونه دوم از جدول ۲، پیامد شاخص ۱۲-CTRP در مقایسه تمرین مقاومتی در برابر کنترل اندازه‌اثری برابر با ۰/۶۲ داشت، اما آزمون گزارش‌شده از سوی مقاله از نظر آماری معنادار نبود، با مقدار احتمال ۰/۲۵۲ (جدول ۲). این وضعیت، که در نگاه اول متناقض به نظر می‌رسد، نمونه کلاسیک غیرمعنادار، ولی مهم از نظر عملی می‌باشد (۵، ۶، ۱۶). بدان معنی، جایی که شدت اثر بالقوه قابل توجه است، اما عدم قطعیت (که معمولاً با بازه اطمینان بهتر دیده می‌شود) و/یا توان آماری محدود، اجازه نمی‌دهد پژوهشگر با قاطعیت به برچسب معنادار تکیه کند (۱، ۷، ۲۱). همین نمونه، یک پیام آموزشی روشن دارد: غیرمعناداری به‌خودی خود مترادف نبود اثر نیست،

بلکه اغلب یعنی داده‌ها هنوز آن قدر قوی/دقیق نیستند، که اثر را با اطمینان از صفر جدا کنند (۲۱، ۲۲، ۳۷). آنچه این یافته‌ها را برای آموزش روش‌مند ارزشمندتر می‌سازد، آن است که می‌توان آن‌ها را در پرتو توصیه‌های روش‌شناختی اثرگذار و فرایزوهش‌های اخیر، که بر فاصله گرفتن از روایت صفر/یک و حرکت به سوی گزارش برآورد اثر همراه با عدم قطعیت تأکید دارند، خوانش و بازتفسیر کرد (۲، ۲۱، ۵۱). در همین راستا، مطالعات فرایزوهشی حوزه پزشکی ورزشی و علوم تمرین نشان داده‌اند که شفافیت و رویه‌های گزارش‌دهی قابل بازبینی هنوز در همه پژوهش‌ها به سطح مطلوب نرسیده و سوگیری به نفع نتایج معنادار

بحث و نتیجه گیری

در ادبیات روش‌شناسی معاصر، معناداری آماری بیشتر به عنوان شاخصی برای سنجش میزان ناسازگاری داده‌ها با فرض صفر تعریف می‌شود و به‌تنهایی بیانگر اندازه یا اهمیت عملی یک اثر نیست (۲۱، ۳۱، ۳۲). از این رو، تفسیر معناداری آماری بدون توجه به زمینه فیزیولوژیک و ملاحظات تصمیم‌گیری تمرینی یا بالینی می‌تواند به برداشت‌های محدود یا نادقیق در تفسیر نتایج پژوهش‌های کاربردی منجر شود (۸، ۱۹). بیانیه انجمن آمار آمریکا و فراخوان‌های بعدی جامعه آمار، با صراحت یادآور شده‌اند که مقدار احتمال نه اندازه اثر را نشان می‌دهد، و نه به‌تنهایی معیار اهمیت نتیجه است (۲، ۷). بنابراین، داوری علمی نباید صرفاً بر عبور یا عدم عبور از آستانه‌های قراردادی مانند ۰/۰۵ بنا شود، بلکه باید به برآورد اثر و عدم قطعیت پیرامون آن تکیه کند (۱۹-۱۷، ۲۱).

در ادبیات تخصصی علوم ورزشی نیز مطرح شده است که اتکای صرف به معناداری یا غیرمعناداری آماری، بدون توجه به شاخص‌هایی نظیر اندازه اثر، می‌تواند به تفسیرهای نادقیق از بزرگی اثر منجر شود (۱۵، ۱۹، ۲۵). برای نمونه، در یک ممیزی از ارائه‌های یک کنفرانس بین‌المللی علوم ورزشی (۷۷)، بخش بزرگی از نتایجی که با برچسب معنادار ارائه شده بودند، هنگام ارزیابی بر پایه فرضیه‌های حداقل اثر/هم‌ارزی و حدود عدم قطعیت، از نظر بزرگی اثر قطعاً مهم نبودند و در مقابل، نتایج غیرمعنادار نیز به‌ندرت شواهد کافی برای بی‌اهمیت بودن فراهم می‌کردند (۲۷، ۷۷). این مشاهده نشان می‌دهد که حتی در محیط‌های پژوهشی

می‌تواند با مسئله بازپذیری گره بخورد (۳۹، ۴۰، ۴۴). همچنین، بررسی‌های جدید در حوزه‌های نزدیک به گرایش فیزیولوژی ورزشی، از جمله کارآزمایی‌های مداخلات فیزیوتراپی، نشان می‌دهد گزارش فاصله اطمینان و برخی مؤلفه‌های کلیدی گزارش‌دهی هنوز کامل و یکنواخت نیست (۲۷، ۷۳). علاوه بر آن، تحلیل‌های مربوط به نشریات برجسته سلامت و تندرستی نیز حکایت از آن دارند که گزارش اندازه اثر و شاخص‌های عدم قطعیت در عمل می‌تواند کمتر از حد انتظار باشد (۷۱، ۷۲). بنابراین، مسئله کم‌رنگ بودن زبان شدت اثر و عدم قطعیت صرفاً یک چالش محلی نیست. با این تفاوت که مسیر برون‌رفت آن در ادبیات روش‌شناسی روشن‌تر شده است: گزارش دقیق برآوردها، نشان دادن دامنه عدم قطعیت، و پرهیز از تفسیرهای مطلق و آستانه‌محور (۱۷، ۲۱، ۵۱).

در مجموع، یافته‌های تحقیق حاضر نشان می‌دهد که معناداری آماری در گزارش‌دهی بسیاری از مقالات داخلی حضوری پررنگ دارد، اما گزارش منظم اندازه اثر و فاصله اطمینان/عدم قطعیت کمتر وارد متن شده است، و درست در همین فاصله است که احتمال لغزش به سوی تفسیرهای صفر/یک، یا فروکاستن غیرمعناداری به بی‌اثری، افزایش می‌یابد (۱۷، ۵۱). این یافته‌ها زمینه لازم را فراهم می‌کند تا در ادامه مقاله، نسبت‌های نهایی چارچوب ۲×۲ و نیز میزان قابلیت ارائه عدم قطعیت در کل نمونه، به‌صورت نظام‌مند و قابل ممیزی تکمیل شود. به‌ویژه از آن‌رو که هر عدد، همراه با مسیر ردیابی دقیق صفحه/جدول ثبت شده و امکان بازبینی مستقل را حفظ می‌کند (۳۹، ۴۰، ۵۰).

بین‌المللی، روایت صفر/یک می‌تواند از واقعیت برآوردی فاصله بگیرد و ضرورت گذار به گزارش هم‌زمان اندازه اثر و عدم قطعیت را تقویت می‌کند (۸، ۲۷، ۵۱).

یافته‌های فراپژوهش حاضر، که بر پایه استخراج نظام‌مند از متن کامل مقالات گرایش فیزیولوژی ورزشی ایران در بازه ۲۰۱۹ تا ۲۰۲۵ و با ثبت ردپای ممیزی برای هر عدد انجام شد، تصویری کم‌ابهام از الگوی گزارش‌دهی ارائه می‌کند: مقدار احتمال در بخش عمده‌ای از مقالات حضور پررنگ دارد، اما زبان دقیق‌تر شدت اثر و عدم قطعیت، یعنی اندازه اثر و فاصله اطمینان/شاخص‌های عدم قطعیت، کم‌فروغ است. در بدنه قابل ممیزی این مطالعه، گزارش مقدار احتمال در ۴۲ مقاله (۹۳،۳٪) دیده شد و گزارش مقدار احتمال دقیق در ۲۶ مقاله (۵۷،۸٪) ثبت گردید، در حالی که گزارش اندازه اثر و گزارش فاصله اطمینان هر کدام تنها در ۲ مقاله (۴،۴٪) مشاهده شد. این ناهم‌ترازی، از منظر مفهومی همان نقطه‌ای است که ادبیات روش‌شناسی آن را منشأ قطعیت کاذب می‌داند (۷، ۱۷، ۲۱): وقتی یک عدد کوچک، نقش داور نهایی را بر عهده می‌گیرد و روایت علمی به حکم قبول/رد فروکاسته می‌شود، دو خطای کلاسیک به‌سادگی رخ می‌دهد: معناداری = اهمیت و غیرمعناداری = بی‌اثری (۲۰-۲۲).

بخش یافته‌های برآوردمحور در این مقاله، هرچند در این مرحله محدود به مقایسه‌هایی بود که پیامد شاخص روشن و داده عددی کافی پس‌آزمون داشتند، به‌صورت آموزشی نشان می‌دهد چرا اتکای صرف به معیارهای آستانه‌ای معناداری می‌تواند گمراه‌کننده

باشد (۸، ۱۹، ۲۱). در نمونه‌ای، اثر از نظر آماری در محدوده معناداری قرار می‌گیرد، اما بزرگی برآورد در مرز اثرهای کوچک می‌باشد. در نمونه‌ای دیگر، برآورد اثر از نظر عملی بزرگ‌تر به نظر می‌رسد، اما نتیجه از منظر آماری غیرمعنادار گزارش شده است. این هم‌نشینی ظاهراً متناقض، دقیقاً همان پیام مرکزی رویکرد برآوردمحور را زنده می‌کند: غیرمعناداری می‌تواند محصول توان ناکافی، پراکندگی بالا، یا ناهمگنی پاسخ‌ها باشد، نه نشانه قطعی نبود اثر (۱۹، ۲۱، ۳۰)، و معناداری نیز می‌تواند در کنار اثری کوچک رخ دهد، بی‌آن‌که الزاماً به معنای ارزش عملی قابل اعتنا باشد (۵، ۲۰، ۲۱).

از منظر چرایی شکل‌گیری این الگو در بخشی از پژوهش‌های داخلی، چند سازوکار هم‌افزا قابل طرح است. فشار انتشار و گرایش نظام‌مند به نتایج مثبت می‌تواند زمینه را برای رفتارهای پژوهشی مسئله‌دار فراهم کند (۳۹، ۴۰، ۴۴). از جمله آزمون‌های متعدد، انتخاب گزینشی پیامدها، و انعطاف پنهان در تحلیل و گزارش‌نویسی که در ادبیات روش‌شناسی با مفاهیمی مانند باغ راه‌های منشعب و پی‌هکینگ توصیف شده‌اند (۲۳، ۲۴). همچنین مسئله توان آماری در مطالعات تمرینی که اغلب با نمونه‌های کوچک و متغیرهای فیزیولوژیک پرنوسان سروکار دارند، می‌تواند سبب شود اثرهای بالقوه مهم، در آزمون‌های آستانه‌ای به‌عنوان غیرمعنادار ظاهر شوند (۱۹، ۲۱، ۳۰). در حالی‌که، گزارش فاصله اطمینان می‌توانست نشان دهد داده‌ها با چه دامنه‌ای از اثرها سازگارترند و آیا اثر مهم واقعاً از دایره امکان بیرون است یا نه (۱، ۲۱، ۷۴). در کنار

این‌ها، چندپیمادی بودن پژوهش‌های ورزشی و تفاوت در نوع آزمون‌های گزارش‌شده (درون‌گروهی در برابر بین‌گروهی یا اثرهای تعاملی) نیز تفسیر را حساس‌تر می‌کند (۵۲، ۵۳، ۵۵)، و اهمیت ثبت نوع آزمون در کنار مقدار احتمال را برجسته می‌سازد (۴۵، ۴۶).

در گرایش فیزیولوژی ورزشی تأکید بر گزارش دقیق مؤلفه‌های آماری صرفاً یک ظرافت دانشگاهی نیست، بلکه بخشی از اخلاق حرفه‌ای تولید شواهد است (۴۹). بدن انسان با پاسخ‌های ناهمگون، خطای اندازه‌گیری، سازگاری‌های زمان‌مند و تفاوت‌های بین‌فردی تعریف می‌شود. بنابراین، گزارش اندازه اثر و عدم قطعیت برای مربی، فیزیولوژیست ورزشی، متخصص توان‌بخشی، و حتی سیاست‌گذار سلامت، همان ابزاری است که کمک می‌کند میان تغییر کوچک اما قابل تکرار، تغییر بزرگ اما نامطمئن، و تغییر ظاهراً معنادار، اما کم‌ارزش عملی، تمایز گذاشته شود (۳، ۵، ۱۶). وقتی گزارش آماری به یک آستانه تقلیل یابد، تصمیم‌گیری حرفه‌ای ناچار می‌شود روی زمینی با خطوط کم‌رنگ حرکت نماید. اما وقتی برآورد و عدم قطعیت کنار هم می‌آیند، نقشه تصمیم‌گیری از حالت سیاه/سفید به طیفی قابل دفاع تبدیل می‌شود (۱۷، ۲۱، ۲۶).

بر این اساس پیشنهاد‌های اجرایی این مطالعه می‌تواند روشن و کم‌هزینه باشد. برای نویسندگان، گزارش مقدار احتمال دقیق به جای عبارت‌های آستانه‌ای، و گزارش اندازه اثر همراه با تعریف روشن نوع آن و ارائه فاصله اطمینان ۹۵ درصد (یا دست‌کم شاخص‌های عدم قطعیت) باید به حداقل استاندارد

گزارش‌دهی بدل شود (۴۲، ۵۰، ۵۱). با این حال، هرگاه مقدار احتمال دقیق در مقاله ارائه نشده باشد، صورت‌بندی آستانه‌ای باید به‌گونه‌ای گزارش شود که از دقت‌نمایی نادرست پرهیز گردد؛ بنابراین به جای نوشتن $p = 0/0000$ ، قالب‌های استاندارد $p < 0,01$ یا $p < 0,05$ یا $p < 0,001$ توصیه می‌شود (۷، ۵۱). برای داوران، ارزیابی ادعاها باید از آیا معنادار است؟ به سمت اندازه اثر چقدر است و عدم قطعیت آن چه می‌گوید؟ جابه‌جا شود (۲، ۱۹، ۲۱)، و برای سردبیران، می‌توان با یک چک‌فهرست کوتاه گزارش‌دهی (شامل نوع آزمون اصلی، مقدار احتمال دقیق، اندازه اثر، فاصله اطمینان، و مسیر ممیزی اعداد کلیدی) کیفیت را از سطح توصیه اخلاقی به سطح قاعده حرفه‌ای نزدیک کرد (۳۹، ۴۰، ۵۰). در این مسیر، هم‌راستا شدن با راهنماهای گزارش‌دهی تثبیت‌شده و تأکید بر شفافیت و بازبینی‌پذیری، می‌تواند به کاهش زبان قطعی و تقویت زبان سازگاری داده‌ها با دامنه‌ای از اثرها بینجامد، زبانی که هم دقیق‌تر و هم علمی‌تر خواهد بود.

با وجود این، مطالعه حاضر محدودیت‌هایی دارد که باید در تفسیر نتایج دیده شود. نخست، بخشی از مقالات به دلیل فقدان داده عددی کافی یا ارائه داده‌ها به‌شکلی که بازسازی برآورد، و عدم قطعیت را دشوار می‌کرد، امکان محاسبه اندازه اثر و فاصله اطمینان را به‌طور کامل فراهم نداشتند (۴۳، ۶۷، ۶۸). بنابراین، در برخی موارد فقر گزارش‌دهی خود به مانعی برای تکمیل برآورد محور تبدیل شد، نه آن‌که لزوماً نشان‌دهنده نبود اثر باشد. دوم، ناهمگنی پیامدها و

تفاوت در آزمون‌های گزارش شده (به‌ویژه اتکا به آزمون‌های درون‌گروهی) می‌تواند نسبت‌های معنادار/غیرمعنادار را تحت تأثیر قرار دهد، و نشان می‌دهد چرا ثبت دقیق نوع آزمون و مسیر ممیزی برای هر عدد ضروری می‌باشد (۵۲، ۵۳، ۵۵). سوم، انتخاب پیامد شاخص، هرچند برای مقایسه‌پذیری ناگزیر، می‌تواند بخشی از واقعیت چندپیامدی پژوهش‌های ورزشی را ساده‌سازی کند، و باید با احتیاط خوانده شود (۴۵، ۴۶).

در جمع‌بندی، این فرایند پژوهش نشان می‌دهد که در بخشی از پژوهش‌های فیزیولوژی ورزشی در ایران بین سال‌های ۲۰۱۹ تا ۲۰۲۵، گزارش‌دهی مبتنی بر مقدار احتمال بسیار شایع‌تر از گزارش‌دهی مبتنی بر اندازه اثر و عدم قطعیت است (۲، ۷، ۱۹)، در حالی که نمونه‌های واقعی نشان می‌دهند ممکن است اثری از نظر عملی قابل توجه باشد، اما به دلیل توان یا پراکندگی، غیرمعنادار گزارش شود (۱۷، ۲۱، ۲۶). البته، عکس این مورد هم ممکن است نتیجه‌ای معنادار باشد، اما بزرگی اثر آن در مرزهای اهمیت کاربردی قرار گیرد. پیام عملی مطالعه حاضر آن است که گزارش‌دهی آماری در گرایش فیزیولوژی ورزشی باید از آستانه‌محوری به سمت برآورد محوری شفاف و قابل ممیزی حرکت نماید (۱۸، ۵۰، ۵۱)، حرکتی که هم آموزش/پژوهش دانشگاهی را دقیق‌تر می‌کند، و هم تصمیم‌گیری تمرینی/بالینی را کم‌خطا تر خواهد کرد.

این مطالعه با رویکرد فرایند پژوهش، الگوی گزارش‌دهی مؤلفه‌های آماری را در مقالات گرایش فیزیولوژی ورزشی منتشر شده در ایران (۲۰۱۹ تا

۲۰۲۵) بررسی کرد، و با ثبت ردپای ممیزی برای هر عدد، امکان بازبینی مستقل را حفظ نمود. یافته‌های توصیفی نشان داد گزارش مقدار احتمال در اکثر مقالات دیده می‌شود، اما گزارش اندازه اثر و فاصله اطمینان/عدم قطعیت در درصد بسیار اندکی از مقالات حضور دارد. در بخش برآورد محور، نمونه‌های قابل محاسبه نشان دادند که هم‌راستایی معناداری آماری با اهمیت عملی همواره بدیهی نیست. یک نتیجه ممکن است معنادار باشد اما از نظر عملی کوچک تلقی شود، یا غیرمعنادار باشد، اما برآورد اثر از منظر عملی قابل توجه به نظر برسد. بر پایه این شواهد، توصیه اجرایی مطالعه آن است که گزارش نتایج در فیزیولوژی ورزشی باید به‌صورت حداقلی شامل مقدار احتمال دقیق، تعریف روشن نوع اندازه اثر، و ارائه فاصله اطمینان یا شاخص‌های عدم قطعیت باشد. همچنین، برای ارتقای کیفیت گزارش‌دهی، در مواردی که مقدار احتمال به‌صورت دقیق ارائه نشده است، استفاده از قالب‌های آستانه‌ای استاندارد ($p < 0.01$ یا $p < 0.05$ یا $p < 0.001$) و پرهیز از نگارش‌های نادرست، مانند $p = 0.0000$ ضروری می‌باشد. زیرا، در این گرایش، تصمیم‌های تمرینی و بالینی بر شانه عددها استوار می‌شود، و عدد نامناسب، بیش از آن که راهنما باشد، می‌تواند منشأ برداشت‌های ساده‌انگارانه گردد.

تشکر و قدردانی

نویسندگان، از تمامی صاحب‌نظران و همکارانی که با ارائه دیدگاه‌ها و بحث‌های علمی در شکل‌گیری ایده اولیه این پژوهش نقش داشته‌اند، صمیمانه قدردانی می‌نمایند.

References

1. Rovetta A, Piretta L, Mansournia MA. p-Values and confidence intervals as compatibility measures: guidelines for interpreting statistical studies in clinical research. *The Lancet Regional Health - Southeast Asia*. 2025;33:100534.
2. Wasserstein RL, Schirm AL, Lazar NA. Moving to a World Beyond “ $p < 0.05$ ”. *The American Statistician*. 2019;73(sup1):1-19.
3. Drinkwater E. Applications of confidence limits and effect sizes in sport research. 2008.
4. Buchheit M. The Numbers Will Love You Back in Return—I Promise. *International Journal of Sports Physiology and Performance*. 2016;11(4):551-4.
5. Batterham AM, Hopkins WG. Making meaningful inferences about magnitudes. *Int J Sports Physiol Perform*. 2006;1(1):50-7.
6. Zemková E. Significantly and Practically Meaningful Differences in Balance Research: P Values and/or Effect Sizes? *Sports Medicine*. 2014;44(7):879-86.
7. Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician*. 2016;70(2):129-33.
8. Bernards JR, Sato K, Haff GG, Bazyler CD. Current research and statistical practices in sport science and a need for change. *Sports*. 2017;5(4):87.
9. Sedaghat H, Taghian F, jalali dehkordi K, Naghizadeh H. Effect of High-Intensity Functional Training Protocol Combined with Ginkgo Biloba SuPPlementation on the Antioxidant Status of Male Firefighters in the OPerational Division of Yazd City. *The Journal of Shahid Sadoughi University of Medical Sciences*. 2025;33(4):8946-60.
10. Mehdi-pour F, Damirchi A, Saidie P, Saberi A. Impact of Three Home-Based Exercise Programs on Metabolic Status and Fat Percent in Women with Multiple Sclerosis. *Metabolism and Exercise*. 2024;14(1):42-58.
11. Tavakoli M, Marandi SM, Kargar Fard M, Pakrad B, Dehghan A. Comparison of trained and untrained girls' blood buffering capacity response to three types of recovery during Repeated High-Intensity endurance test. *Journal of Sport and Exercise Physiology*. 2021;14(2):55-66.
12. Bok D, Foster C. Applicability of Field Aerobic Fitness Tests in Soccer: Which One to Choose? *Journal of functional morphology and kinesiology*. 2021;6(3).
13. Ahmadizad S, Purkiani M, Taghizadeh MR, Rahmani H, Emami F. Investigation and comparison of anthropometric, somatotype and physical fitness characteristics of the twelfth students' Olympiad medal winners with national champions. *Journal of Sport and Exercise Physiology*. 2021;14(2):87-100.
14. Masanovic B, Popovic S, Bjelica D. Comparative study of anthropometric measurement and body composition between basketball players from different competitive levels: elite and sub-elite. *Pedagogics, psychology, medical-biological problems of physical training and sports*. 2019(4):176-81.
15. Caldwell A, Vigotsky AD. A case against default effect sizes in sport and exercise science. *PeerJ*. 2020;8:e10314.
16. Riemann BL, Lininger M. Statistical Primer for Athletic Trainers: The Difference Between Statistical and Clinical Meaningfulness. *Journal of athletic training*. 2015;50(12):1223-5.
17. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature*. 2019;567(7748):305-7.

18. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon Statistical Significance. *The American Statistician*. 2019;73(sup1):235-45.
19. Hopkins WG, Marshall SW, Batterham AM, Hanin J. Progressive statistics for studies in sports medicine and exercise science. *Medicine and science in sports and exercise*. 2009;41(1):3-13.
20. Cohen J. The earth is round ($p < .05$). *American psychologist*. 1994;49(12):997.
21. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology*. 2016;31(4):337-50.
22. Altman DG, Bland JM. Statistics notes: Absence of evidence is not evidence of absence. *BMJ*. 1995;311(7003):485.
23. Gelman A, Loken E. The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. Department of Statistics, Columbia University. 2013;348(1-17):3.
24. Simmons JP, Nelson LD, Simonsohn U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*. 2011;22(11):1359-66.
25. Harrison AJ, McErlain-Naylor SA, Bradshaw EJ, Dai B, Nunome H, Hughes GTG, et al. Recommendations for statistical analysis involving null hypothesis significance testing. *Sports biomechanics*. 2020;19(5):561-8.
26. Amrhein V, Greenland S, McShane BB. Statistical significance gives bias a free pass. *European journal of clinical investigation*. 2019;49(12):e13176.
27. Hopkins WG. Replacing statistical significance and non-significance with better approaches to sampling uncertainty. *Frontiers in physiology*. 2022;13:962132.
28. Keller M, Lichtenstein E, Roth R, Faude O. Balance Training Under Fatigue: A Randomized Controlled Trial on the Effect of Fatigue on Adaptations to Balance Training. *The Journal of Strength & Conditioning Research*. 2024;38(2):297-305.
29. Poon ET-C, Li H-Y, Little JP, Wong SH-S, Ho RS-T. Efficacy of Interval Training in Improving Body Composition and Adiposity in Apparently Healthy Adults: An Umbrella Review with Meta-Analysis. *Sports Medicine*. 2024;54(11):2817-40.
30. Hopkins WG. Sample Sizes for Magnitude-Based Inferences about Clinical, Practical or Mechanistic Significance: 2746: Board #20 9:AM – 10:AM. *Medicine & Science in Sports & Exercise*. 2006;38(5):S528-S9.
31. Helalizadeh M, Taghizadeh MR, Afshari M, Irani R. Content Analysis of Journal of Sport Physiology. *Sport Physiology*. 2017;9(34):17-34.
32. Arzani A, Mirdar S, Asghari B. Citation and Content Analysis of Journals of Applied Exercise Physiology, Sport Management and Motor Behavior at the University of Mazandaran. *Journal of Applied Exercise Physiology*. 2018;14(27):267-84.
33. Sharifi M, Mirdar S, Sunasgari B. Descriptive Analysis of Articles of Physical Education and Sport Sciences Journals in 1384-1388. *Sport Management Journal*. 2012;4(12):61-85.
34. Earnest CP, Roberts BM, Harnish CR, Kutz JL, Cholewa JM, Johannsen NM. Reporting Characteristics in Sports Nutrition. *Sports*. 2018;6(4):139.
35. Figueiredo P, Marques EA, Backes OP, Gaya AR, Castro FdS. A review of the statistical methods used in sports and exercise sciences PhD theses-a case study in a single post-graduate program. *R Bras Ci Mov*. 2018;26(4):191-200.
35. Boutron I, Altman DG, Hopewell S, Vera-Badillo F, Tannock I, Ravaud P. Impact of spin in the abstracts of articles reporting results of randomized controlled trials in the field of cancer: the SPIIN randomized controlled trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2014;32(36):4120-6.

36. Boutron I, Dutton S, Ravaud P, Altman DG. Reporting and Interpretation of Randomized Controlled Trials With Statistically Nonsignificant Results for Primary Outcomes. *JAMA*. 2010;303(20):2058-64.
38. Chellamuthu G, Muthu S, Damodaran UK, Rangabashyam R. "Only 50% of randomized trials have high level of confidence in arthroscopy and sports medicine"—a spin-based assessment. *Knee Surgery, Sports Traumatology, Arthroscopy*. 2021;29(9):2789-98.
39. Schulz R, Langen G, Prill R, Cassel M, Weissgerber TL. Reporting and transparent research practices in sports medicine and orthopaedic clinical trials: a meta-research study. *BMJ open*. 2022;12(8):e059347.
40. Warmenhoven J, Menaspà P, Borg DN, Vazire S, White N, Sainani K, et al. Sports Metaresearch: An Emerging Discipline of Sport Science and Medicine. *Sports medicine (Auckland, NZ)*. 2025;55(4):845-56.
41. Impellizzeri FM, Meyer T, Wagenpfeil S. Statistical considerations (or recommendations) for publishing in Science and Medicine in Football. *Science and Medicine in Football*. 2019;3(1):1-2.
42. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC medicine*. 2010;8:18.
43. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*. 2013;4:863.
44. Borg DN, Barnett AG, Caldwell AR, White NM, Stewart IB. The bias for statistical significance in sport and exercise medicine. *Journal of science and medicine in sport*. 2023;26(3):164-8.
45. Mayo-Wilson E, Fusco N, Li T, Hong H, Canner JK, Dickersin K. Multiple outcomes and analyses in clinical trials create challenges for interpretation and research synthesis. *Journal of Clinical Epidemiology*. 2017;86:39-50.
46. Butcher NJ, Monsour A, Mew EJ, Chan AW, Moher D, Mayo-Wilson E, et al. Guidelines for Reporting Outcomes in Trial Reports: The CONSORT-Outcomes 2022 Extension. *Jama*. 2022;328(22):2252-64.
47. Akbari Zarradkhaneh S, attari M. Necessity of Measuring and Reporting Effect Size Measures in Psychological Studies. *Rooyesh-e-Ravanshenasi Journal(RRJ)*. 2016;5(1):29-46.
48. Sepidarkish M, Mohammadi Pirouz Z. An Overview of Statistical and Clinical Concepts with the Approach of Hypothesis Testing (P-value). *Journal of Isfahan Medical School*. 2023;41(732):725-35.
49. Guelmami N, Ezzdine Lb, Ghouili H, Trabelsi O, Ben SH, Glenn JM, et al. The ethical compass: establishing ethical guidelines for research practices in sports medicine and exercise science. 2024.
50. Hopewell S, Chan A-W, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 Statement: Updated Guideline for Reporting Randomized Trials. *JAMA*. 2025;333(22):1998-2005.
51. Johnson SL, Stone WJ, Bunn JA, Lyons TS, Navalta JW. New author guidelines in statistical reporting: Embracing an era beyond $p < .05$. *International Journal of Exercise Science*. 2020;13(1):1.
52. O'Connell NS, Dai L, Jiang Y, Speiser JL, Ward R, Wei W, et al. Methods for Analysis of Pre-Post Data in Clinical Research: A Comparison of Five Common Methods. *Journal of biometrics & biostatistics*. 2017;8(1):1-8.
53. Bland JM, Altman DG. Best (but oft forgotten) practices: testing for treatment effects in randomized trials by separate analyses of changes from baseline in each group is a misleading approach. *The American journal of clinical nutrition*. 2015;102(5):991-4.

54. J T, L B, T H, J R, M W, M H. Different ways to estimate treatment effects in randomised controlled trials. *Contemporary Clinical Trials Communications*. 2018;10:80-5.
55. Kypri K. Interpretation of within-group change in randomised trials. *BMC psychiatry*. 2020;20(1):239.
56. Andrade C. Mean Difference, Standardized Mean Difference (SMD), and Their Use in Meta-Analysis: As Simple as It Gets. *The Journal of clinical psychiatry*. 2020;81(5).
57. Parpia S, Morris TP, Phillips MR, Wykoff CC, Steel DH, Thabane L, et al. Sensitivity analysis in clinical trials: three criteria for a valid sensitivity analysis. *Eye*. 2022;36(11):2073-4.
58. Wang Hm, Wang X, Wu X, Yu K, Jia S, Jiang W. Effects of 24-week jumping exercise on bone mineral density and linear growth in children with short stature: a prospective controlled trial. *BMC Sports Science, Medicine and Rehabilitation*. 2025;17(1):334.
59. Lowery M, Oliver SJ, Roberts R, Barwood C, Dunn E, Langham-Walsh E, et al. A Comparison of Training, Injury, Illness, Sleep, Wellbeing and Stress Between Developing Elite and Recreational Athletes. *European journal of sport science*. 2025;25(12):e70093.
60. Mowbray FI, Manlongat D, Shukla M. Sensitivity Analysis: A Method to Promote Certainty and Transparency in Nursing and Health Research. *The Canadian journal of nursing research = Revue canadienne de recherche en sciences infirmieres*. 2022;54(4):371-6.
61. Tegegne TK, Rawstorn JC, Nourse RA, Kibret KT, Ahmed KY, Maddison R. Effects of exercise-based cardiac rehabilitation delivery modes on exercise capacity and health-related quality of life in heart failure: a systematic review and network meta-analysis. *Open heart*. 2022;9(1).
62. Schünemann HJ, Neumann I, Hultcrantz M, Brignardello-Petersen R, Zeng L, Murad MH, et al. GRADE guidance 35: update on rating imprecision for assessing contextualized certainty of evidence and making decisions. *J Clin Epidemiol*. 2022;150:225-42.
63. De Smet S, Vandecruys M, De Beir J, Leunis S, Wyngaert KV, Renier M, et al. Home-based exercise and physical activity intervention after kidney transplantation: impact of exercise intensity (PHOENIX-Kidney). Protocol for a multicentre randomized controlled trial. *Clinical Kidney Journal*. 2025;18(11).
64. Blackwell JEM, Doleman B, Herrod PJJ, Ricketts S, Phillips BE, Lund JN, et al. Short-Term (<8 wk) High-Intensity Interval Training in Diseased Cohorts. *Medicine and science in sports and exercise*. 2018;50(9):1740-9.
65. Lin L, Aloe AM. Evaluation of various estimators for standardized mean difference in meta-analysis. *Statistics in medicine*. 2021;40(2):403-26.
66. Lee DK. Alternatives to P value: confidence interval and effect size. *Korean journal of anesthesiology*. 2016;69(6):555-62.
67. Gallardo-Gómez D, Richardson R, Dwan K. Standardized mean differences in meta-analysis: A tutorial. *Cochrane evidence synthesis and methods*. 2024;2(3):e12047.
68. Higgins J, Li T, Deeks JJ. Chapter 6: Choosing effect measures and computing estimates of effect 2024 [Available from: <https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current/chapter-06>].
69. Ayooobi F, Rahmani MR, Assar S, Jalalpour S, Rezaeian M. The Consort (Consolidated Standards Of Reporting Trials). *Journal of Rafsanjan University of Medical Sciences*. 2017;15(10):977-94.
70. Hadizadeh-Talasaz Z, Larki M, Latifnejad Roudsari R. Assessing the quality of published clinical trials regarding the effect of conservative methods in women with pelvic floor disorders using CONSORT tool. *The Iranian Journal of Obstetrics, Gynecology and Infertility*. 2023;26(3):109-22.
71. Chu B, Liu M, Leas EC, Althouse BM, Ayers JW. Effect size reporting among prominent health journals: a case study of odds ratios. *BMJ evidence-based medicine*. 2020;26(4):184.

72. Karadaghy OA, Hong H, Scott-Wittenborn N, Sinha P, Suko J, Tait S, et al. Reporting of Effect Size and Confidence Intervals in JAMA Otolaryngology-Head & Neck Surgery. JAMA otolaryngology-- head & neck surgery. 2017;143(11):1075-80.
73. Hernando DF, Elkins M, Freire APCF. Reporting of confidence intervals, achievement of intended sample size, and adjustment for multiple primary outcomes in randomised trials of physical therapy interventions: an analysis of 100 representatively sampled trials. Brazilian Journal of Physical Therapy. 2024;28(3):101079.
74. Fernainy P, Sourial N. Leveraging confidence intervals to inform the clinical significance of interventions. Family practice. 2022;39(6):1196-8.
75. Ampomah K, Amano S, Wages NP, Volz L, Clift R, Ludin AFM, et al. Blood flow–restricted exercise does not induce a cross-transfer of effect: a randomized controlled trial. Medicine and science in sports and exercise. 2019;51(9):1817.
76. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. bmj. 2021;372.
77. Hopkins WG. Misleading Conclusions Based on Statistical Significance and Non-significance at a Recent International Conference. Sportscience. 2021;25:1-10.