



Brazilian Journal of Physical Therapy

<https://www.journals.elsevier.com/brazilian-journal-of-physical-therapy>

ORIGINAL RESEARCH

Single-group studies in leading physical therapy journals commonly make inappropriate conclusions regarding treatment effect modifiers. A systematic review



Tayla J. Douglas¹, Elin Rosenberg Čemažar, Natasha C. Pocovi², Mark J. Hancock^{3,*}

Department of Health Sciences, Macquarie University, Sydney, NSW, Australia

Received 8 July 2022; received in revised form 6 April 2023; accepted 15 June 2023
Available online 22 June 2023

KEYWORDS

Baseline characteristics;
Predictors of response;
Systematic review;
Treatment effect modifiers;
Treatment outcome;
Treatment response

Abstract

Background: Characteristics that identify patients who respond differently to certain interventions are called treatment effect modifiers. Some studies inappropriately report the presence of treatment effect modifiers without adequate study designs.

Objectives: To evaluate what proportion of single-group studies published in leading physical therapy journals inappropriately report treatment effect modifiers, and to assess whether the proportion varies over time or between journals.

Methods: A systematic review was conducted of studies published in eight leading physical therapy journals since 2000. Eligible studies were single-group studies (e.g., cohort study or secondary analysis of treatment arm of randomised controlled trial) that investigated any condition, treatment or outcome. Studies that suggested participants with certain baseline characteristics responded better/or worse to the treatment, were considered to have reported inappropriately. Studies reporting that participants with certain baseline characteristics had improved outcomes but did not state it was due to the treatment were considered to have reported appropriately. The proportion of inappropriate reporting was compared over time and between journals.

Results: Of the 145 included studies, 73 (50.3%) were categorised as inappropriately reporting treatment effect modifiers. The proportion of inappropriate reporting was highest in the most recent period, 2018 – 2022 (59.6%) and 2006 – 2011 (55.6%). The proportion of inappropriate reporting varied substantially between journals from 0% (Journal of Physiotherapy) to 91.7% (Journal of Neurologic Physical Therapy).

Conclusions: A large proportion (50.3%) of single-arm studies in leading physical therapy journals inappropriately report treatment effect modifiers. This inappropriate reporting risks misleading clinicians when selecting interventions for individual patients.

* Corresponding author at: Department of Health Sciences, Macquarie University, Ground Floor 75 Talavera Rd, Sydney, NSW 2109, Australia.
E-mail: mark.hancock@mq.edu.au (M.J. Hancock).

¹ @douglas_tayla

² @PocoviTash

³ @MarkHancockPT

<https://doi.org/10.1016/j.bjpt.2023.100520>

1413-3555/© 2023 The Author(s). Published by Elsevier España, S.L.U. on behalf of Associação Brasileira de Pesquisa e Pós-Graduação em Fisioterapia. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Introduction

Physical therapists utilize a range of interventions to treat their patients. For example, the management of spinal pain may encompass patient education, various exercises, manual therapy, electrothermal techniques, and more.¹ To optimize patient outcomes, physical therapists aim to tailor treatment to suit the patient and their needs. However, this can be challenging as there is limited evidence concerning which patient characteristics predict the response to a specific treatment.^{2,3} Such characteristics are also known as treatment effect modifiers.⁴ Treatment effect modifiers have the potential to improve patient care as physical therapists can provide tailored treatments that are likely to be most beneficial for that individual.⁵

Identifying treatment effect modifiers is recognised as a research priority in physical therapy research and conditions commonly managed by physical therapists such as low back pain.^{6,7} In recent decades, there has been a surge of research exploring effect modifiers. Studies have found that approximately 60% of randomised controlled trials (RCTs) in medical journals explore effect modifiers.^{8,9} An early and well known example from the physical therapy literature is the clinical prediction rule for identifying responders to spinal manipulation by Childs et al.¹⁰ This high quality RCT reported that a subgroup of patients with low back pain who meet the criteria for a clinical prediction rule (derived from baseline characteristics) benefit more from spinal manipulation therapy than those who do not.¹⁰ However, the credibility of many other subgroup (effect modifier) studies has been widely criticised.^{2,3} A good understanding of the appropriate design and reporting of studies investigating effect modifiers is important for both clinicians and researchers.

Without a suitable research design and analysis, researchers can make inappropriate conclusions about effect modifiers. RCTs with well-conducted analyses are needed to identify effect modifiers.⁴ Specifically, studies need tests of interaction, comparing the effect size of one investigated characteristic to another.⁴ Cohort designs offer only single-group data, and without a control group, they cannot estimate the treatment effect size or distinguish effect modifiers from simple predictors of outcome.⁴ Despite this, some cohort studies inappropriately draw conclusions about effect modifiers.¹¹ For example, a 2019 cohort study¹² concludes that four baseline characteristics predict a positive response to motor control training in patients with low back pain. Physical therapists seeking the best evidence to guide care may be misled by claims of effect modification based on inadequate study designs and analyses, and consequently make inappropriate decisions when selecting what they believe is the best intervention for an individual patient.^{2,4} For that reason, inappropriate reporting may be a significant issue that impairs clinical practice and patient care.

The prevalence of inappropriate reporting of treatment effect modifiers remains unknown. It is also unclear whether

inappropriate reporting varies across journals or has changed over time. The present study focused on cohort studies and secondary analyses of RCTs that report findings within a single-treatment group, i.e. do not test for interaction. Such a study should only relay characteristics associated with outcome and prognostic factors. However, we hypothesized that some studies would imply that patient characteristics can predict the response to a specific treatment.

Therefore, the aim of this study was to evaluate what proportion of single-group studies published in leading physical therapy journals inappropriately report treatment effect modifiers, and to assess whether the proportion varies over time or between journals.

Methods

The conduct and reporting of this systematic review are based on the preferred reporting items for systematic reviews and meta-analysis (PRISMA) 2020 guidelines.¹³ The review was prospectively registered on the PROSPERO database (CRD42022304356).

Identification and selection of studies

The investigated journals were all members of the International Society of Physiotherapy Journal Editors (ISPJE). The top eight were selected based on the 2020 journal impact factor (JIF) – as indicated by the Journal Citation Reports database (Clarivate analytics®). The selected journals and their corresponding journal impact factor are Journal of Physiotherapy (7.000), Journal of Orthopaedics & Sports Physical Therapy (4.751), Journal of Neurologic Physical Therapy (3.649), Journal of Geriatric Physical Therapy (3.381), Brazilian Journal of Physical Therapy (3.377), Physiotherapy (3.358), Physical Therapy and Rehabilitation Journal (3.021) (previously known as Physical Therapy), and Musculoskeletal Science and Practice (2.520) (previously known as Manual Therapy).

A search was executed using Ovid Medline. Additional databases were unnecessary as Ovid Medline had access to all relevant articles of the included journals. We developed the search strategy based on pilot testing to limit studies to those potentially relevant to the inclusion criteria described below. The final search (Supplementary Material – Table S1) contained the terms ‘cohort’, ‘randomised control trial’, and ‘predict’. We also restricted results by year (2000 – current).

To be eligible for inclusion, studies had to meet the following criteria: (1) a published cohort study or RCT; (2) a sample size of 10 or more; (3) participants received a uniform treatment; (4) report a relationship between one or more baseline patient characteristics and outcome in a single-treatment group, i.e. do not compare to a control group or test for interaction (in the case of an RCT design); (5) published in one of the top eight physical therapy journals; and (6) published from 2000 up to the search date. We defined

uniform treatment as any treatment that was consistent between participants. For example, all participants received ‘manual therapy’. We also considered broad interventions such as ‘physical therapy’ or ‘multidisciplinary care’ uniform. Eligibility was not restricted by participant demographics, diagnosis, type of intervention, or outcome.

Three independent reviewers (TD, ERC, and NP) screened the studies by title and abstract and then by full text. Any disagreements were resolved by discussion, or a by a fourth reviewer (MH) if required.

Data extraction / assessment of study characteristics

A data extraction form was developed and pilot tested by two reviewers. The data extraction was performed independently by two reviewers (TD and ERC). The following data were recorded: title of study; author(s); journal; year published; study design; treatment; condition; outcome(s); baseline characteristic(s); and appropriateness of reporting. Any disagreements were resolved through discussion or by a third reviewer (MH) if required.

Assessment of appropriateness of reporting

Before the review commenced, the reviewers extracting data (TD and ERC) underwent extensive, uniform training on assessing appropriateness to optimize reliability. This included assessing examples and subsequent discussion with an expert in the field (MH). For our primary outcome, studies were deemed appropriate or inappropriate based on how the relationship between baseline characteristics and outcome was reported (see [Table 1](#) for full details). Terms such as ‘respond’, ‘benefit’, and ‘treatment effect’ were considered inappropriate as they imply that the characteristics influenced the treatment response. Readers can review specific examples from studies we considered to inappropriately report effect modification in Supplementary Material –

Table S2. We reviewed reporting throughout the full papers. However, more emphasis was placed on the ‘key sections’ of the paper, such as the title, abstract, and final conclusion, as these sections are more pronounced and contain the principal findings. Studies were deemed inappropriate if they contained one inappropriate term in a ‘key section’ or two or more inappropriate terms elsewhere in the paper. The introduction section was also considered, where authors discuss previous works to justify their research aim.

We expected some studies to use caution despite inappropriately reporting effect modification – for instance, those who acknowledge their findings’ limitations due to the unsuitable study design. Other studies are inconsistent in their reporting, using both appropriate and inappropriate language. Therefore, as a secondary outcome we created a “mildly inappropriate” subcategory to describe these studies. The full criteria used for the assessment of the level of inappropriateness can be found in [Table 1](#).

Study appropriateness was analysed individually by two reviewers (TD and ERC), and a third reviewer (MH) resolved the remaining disagreements. A fourth reviewer was consulted if the primary reviewers disagreed on a study with which the third reviewer was an author.

Quality assessment of the included studies was unnecessary due to the aim of the study.

Data analysis

To determine the proportion of studies that had inappropriate reporting, we used the following equation: the number of inappropriate studies divided by the total number of included studies. The secondary analysis focused on the inappropriate subcategories: the number of mildly inappropriate studies divided by the total number of inappropriate studies. Because each of these calculated proportions were based on all eligible studies in the 8 journals of interest (the full population like a census), rather than a randomly selected sample of the eligible studies, the 95% confidence

Table 1 Assessment of appropriateness of reporting criteria.

Appropriateness category	Reporting style	Examples
Appropriate	Use only association or prognostic language when reporting the relationship between baseline patient characteristics and outcome.	<ul style="list-style-type: none"> • Females are more likely to have lower pain levels. • Anxiety is a risk factor for delayed recovery. • Patients with a body mass index greater than 25 have a poorer prognosis.
Inappropriate	Use language that suggests that patients with specific baseline characteristic/s have a distinct treatment response to the investigated intervention.	<ul style="list-style-type: none"> • Participants over the age of 65 were shown to benefit more from the treatment. • A clinical prediction rule was developed to predict a successful response to spinal manipulation. • Patients with less range of motion demonstrated improved treatment response.
Mildly inappropriate	Same as inappropriate (above) but also uses some cautious language when reporting relationships between baseline patient characteristics and treatment response.	<ul style="list-style-type: none"> • The design of this study cannot establish a cause-and-effect relationship between the treatment and outcome. • The findings of this study should be validated in a randomised controlled trial.

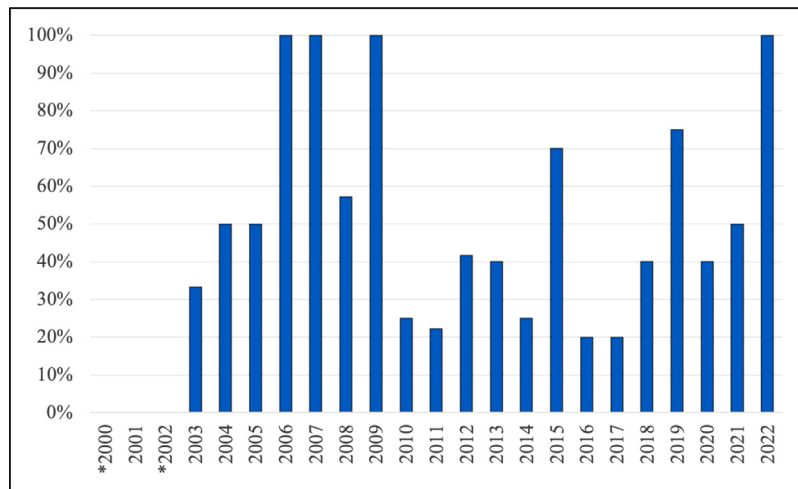


Fig. 1 Proportion of inappropriate reporting by year of publication.
*No included studies were published in these years.

interval (CI) for each proportion was not calculated. CIs are used to describe uncertainty of the population value when a sample is used.

To explore whether inappropriate reporting changes over time or differs between journal we observed the differences in inappropriate reporting for these two variables. We did not run statistical tests, or tests of statistical inference as each population was “complete” as the observed differences between categories were exact and not an estimate. For the variable time, we originally intended to investigate it as a continuous variable but as the relationship with inappropriate reporting clearly violated the linearity assumption (Fig. 1) we created a categorical variable with four roughly equal periods of time (2000 – 2005, 2006 – 2011, 2012 – 2017, 2018 – 2022) to help describe the trends over time.

Results

Flow of studies through the review

A total of 3287 potentially eligible studies were retrieved from the electronic search. 3016 studies were excluded by title and abstract, and another 126 were excluded when reviewed by full text. Therefore, the remaining sample consisted of 145 studies. The flow of studies through the review is summarised in Fig. 2. See supplementary material – Table S3 for the full list of included studies.

Characteristics of the included studies

The majority of included studies were published in Physical Therapy and Rehabilitation Journal and Journal of Orthopaedics & Sports Physical Therapy, 32% and 21%, respectively. The number and proportion of studies in each of the four periods of time (2000 – 2005, 2006 – 2011, 2012 – 2017, 2018 – 2022) are presented in Table 2, with the smallest proportion of included studies in the earliest period, 2000 – 2005, which accounted for only 8% of included studies. The most common treatments investigated were general physical therapy (32%),

rehabilitation (17%), manual therapy (12%), and exercise (11%). The conditions investigated in the studies were mostly musculoskeletal (68%), followed by neurological (14%) and cardiorespiratory (4%). Table 2 summarises the descriptive details of included studies.

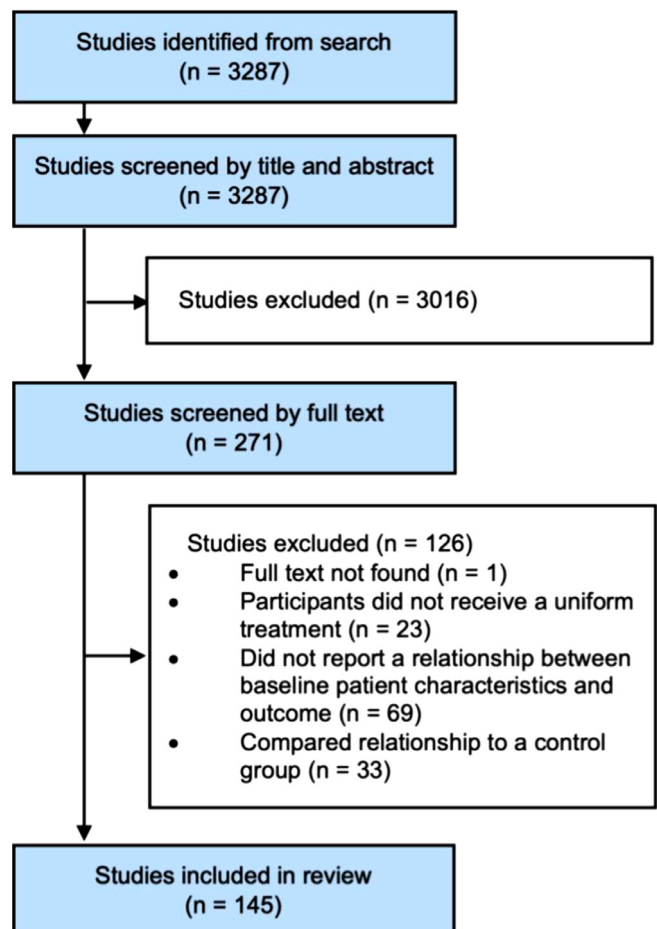


Fig. 2 Flow of studies through the review.

Table 2 Descriptive details of included studies.

Descriptive details of included studies	(n = 145)
Journal, n (%)	
Journal of Physiotherapy	5 (3)
Journal of Orthopaedics & Sports Physical Therapy	31 (21)
Journal of Neurologic Physical Therapy	12 (8)
Journal of Geriatric Physical Therapy	12 (8)
The Brazilian Journal of Physical Therapy	7 (5)
Physiotherapy	8 (6)
Physical Therapy and Rehabilitation Journal (previously Physical Therapy)	47 (32)
Musculoskeletal Science and Practice (previously Manual Therapy)	23 (16)
Year of publication, n (%)	
2000 – 2005	12 (8)
2006 – 2011	36 (25)
2012 – 2017	50 (34)
2018 – 2022	47 (32)
Study design, n (%)	
Cohort study	112 (77)
Secondary analysis of a cohort study	5 (3)
RCT	7 (5)
Secondary analysis of RCT	21 (14)
Participants / Condition, n (%)	
Musculoskeletal	98 (68)
Neurological	20 (14)
Cardiorespiratory	6 (4)
Other	
Paediatric	3 (2)
Geriatric	5 (3)
Hospital inpatient	2 (1)
Amputee	2 (1)
Urogenital	4 (3)
Vision and vestibular disorders	3 (2)
Lymphoedema	1 (0.5)
Fibromyalgia	1 (0.5)
Interventions, n (%)	
Physical therapy	
General physical therapy (or chiropractic)	47 (32)
Rehabilitation	25 (17)
Exercise	16 (11)
Manual therapy	18 (12)
McKenzie method	5 (3)
Taping and bandaging	3 (2)
Surgery	
Total hip replacement	5 (3)
Total knee replacement	7 (5)
ACL reconstruction	1 (0.5)
Meniscal surgery	1 (0.5)
Neck of femur fixation	1 (0.5)
Femoral lengthening surgery	1 (0.5)
Thoracic surgery	3 (2)
Other	
Multidisciplinary care / program	4 (3)
Cognitive behavioural therapy	1 (0.5)
Robotics and gaming technology	3 (2)
Vestibular rehabilitation	1 (0.5)
Optokinetic training	1 (0.5)
Skilled nursing services	1 (0.5)
Dialysis	1 (0.5)

Appropriateness of reporting

From the 145 eligible studies, 73 (50.3%) were categorised as inappropriately reporting treatment effect modifiers. Out of the 73 that were categorised as having inappropriate reporting, 37 (50.7%) were categorised as mildly inappropriate. The appropriateness ratings for all included studies can be found in the supplementary material – Table S3.

The proportion of studies inappropriately reporting treatment effect modifiers for the set time periods is shown in Fig. 3. The proportion in the earliest period was 41.7% which increased to 55.6% in the subsequent period of time (2006–2011). After this the proportion reduced to 40.0% from 2012 to 2017 and then increased to 59.6% in the most recent period (2018–2022).

The proportion of studies that had inappropriate reporting for each of the eight included journals is displayed in Fig. 3. The proportions varied substantially between journals from 0% (0 of 5) for Journal of Physiotherapy to 91.7% (11of12) for Journal of Neurologic Physical Therapy. Musculoskeletal Science and Practice, Physical Therapy and Rehabilitation Journal, and Journal of Geriatric Physical Therapy had quite high rates of inappropriate reporting (69.6, 53.2, and 50.0% respectively), while Brazilian Journal of Physical Therapy had lower rates of inappropriate reporting (14.3%).

Discussion

Main findings

The present study found that 50.3% of single-group studies from the top eight physical therapy journals inappropriately reported treatment effect modifiers. Approximately half of the inappropriately reported studies (50.7%) used some caution suggesting that authors may have been aware that their study design was not suitable for drawing conclusions about effect modifiers. The proportion of inappropriate reporting appears to have varied over time. There was a spike in the 2006 – 2011 period (55.6%) and a more recent spike (59.6%) between 2018 and 2022. The proportion of inappropriate reporting varied substantially between journals from 0% (Journal of Physiotherapy) to 91.7% (Journal of Neurologic Physical Therapy).

Comparison with other studies

There has been limited investigation of how commonly inappropriate reporting of effect modifiers occurs, but a related study was conducted in 2018 by Dahan et al.⁵ They investigated how many cohort studies measured and reported on the association between a baseline characteristic and outcome from 2000 to 2016.⁵ They called this an investigation of heterogeneity of treatment effect (HTE) and found that 59% of cohort studies reported HTE.⁵ While their study investigated the proportion of cohort studies reporting an association between a baseline characteristic and outcome, it did not analyze whether these cohort studies were reporting the association appropriately (i.e., prognosis) or inappropriately (i.e., treatment effect modifiers). This is important because the issue is not that researchers use cohort studies to investigate these associations, but rather how they report this

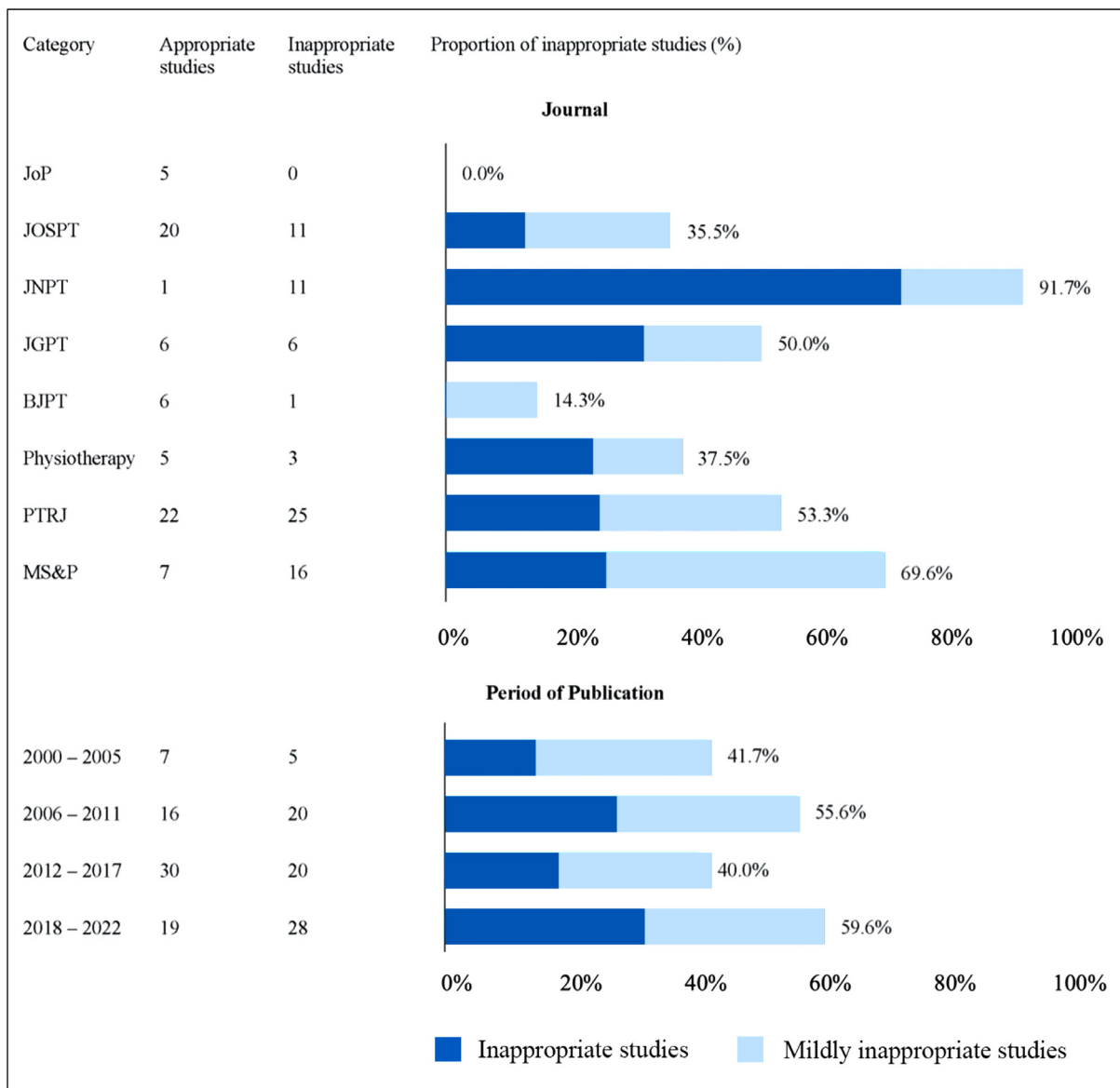


Fig. 3 Inappropriate reporting by year of publication and journal. JoP, Journal of Physiotherapy; JOSPT, Journal of Orthopaedics & Sports Physical Therapy; JNPT, Journal of Neurologic Physical Therapy; JGPT, Journal of Geriatric Physical Therapy; BJPT, Brazilian Journal of Physical Therapy; PTRJ, Physical Therapy and Rehabilitation Journal (previously Physical Therapy); MS&P, Musculoskeletal Science and Practice (previously Manual Therapy).

information. The design of our systematic review incorporates this analysis and investigates the appropriateness of reporting made in cohort studies, specifically focusing on the type of language used by researchers.

The current study does not provide evidence of why inappropriate reporting is common in the physical therapy literature or why it appears to have peaked in the 2006 – 2011 period and again more recently. However, the 2006 – 2011 peak occurred soon after the high profile and highly cited clinical prediction rule for identifying responders to spinal manipulation, published in 2004 in *Annals of Internal Medicine*.¹⁰ This study used a RCT design and drew appropriate conclusions. However, enthusiasm for identifying effect moderators may have contributed to many subsequent studies drawing these conclusions despite not using appropriate

designs or analyses. Following this, several authors reported the problems with using cohort designs to investigate treatment effect modifiers,^{4,11} which may have contributed to the drop in inappropriate reporting in the subsequent period. But, in the most recent period inappropriate reporting has increased again.

Meaning and implications of findings

The present study reveals that inappropriate reporting of effect modification in cohort studies is an ongoing problem in most of the investigated physical therapy journals, as 50.3% of eligible studies used language that implies they have identified a treatment effect modifier. This indicates the need for clinicians to use care when interpreting such

findings to make decisions regarding which interventions to prescribe to an individual patient. Journals should establish clear reporting guidelines or a publishing policy for single-arm cohort studies that investigate baseline patient characteristics. We believe such policies will help mitigate inappropriate reporting in the future, ensuring studies do not mislead clinicians in ways that could compromise patient care. There is an extensive body of literature on how to properly conduct and report studies investigating effect modifiers and avoid inappropriate reporting in single-arm studies.^{2,4,14–16} We encourage interested clinicians, researchers, and journal editors to refer to this literature.

Limitations

The primary limitation of our study is the lack of a previously published tool for assessing appropriateness of reporting of effect modifiers. However, our criteria are based on principles widely reported in the literature reporting on correct conduct and analysis or studies investigating effect modifiers or subgroups.^{2–4,8,11,14,17} Reviewers underwent training and pilot tested the appropriateness criteria on similar studies from other journals prior to extracting data until high levels of agreement were achieved. We assessed agreement between the two raters (prior to meeting to come to consensus) on a subset of 47 included studies and found excellent agreement with a Kappa of 0.90. It is possible that our search criteria may have missed some relevant studies. To assess this, we reviewed all articles (980) published in one of the journals (Journal of Physiotherapy) since 2000 and found no relevant articles missed by our search. Our results are representative of the eight included journals but caution needs to be taken generalizing these findings to all physical therapy journals. The four time-windows were selected to produce relatively equal periods of time since 2000. If different time-windows were selected the results may then differ. The proportion of inappropriate reporting per year is presented in Fig. 1 to help readers judge change over time.

Future research

There are several opportunities for future research in this area. Inappropriate reporting of treatment effect modifiers needs to be formally defined. Literature that aims to define inappropriate reporting will increase the validity of future studies in this area. A definition may also help clarify and build support for the concept, indirectly resulting in less inappropriate reporting. The present study also focused on 8 leading physical therapy journals. We encourage future researchers to investigate inappropriate reporting in other journals covering a breadth of disciplines and journal impact factors. It is not clear if the problem is more or less common in physical therapy journals than journals covering other areas of medical research.

Conclusion

We found a high proportion (50.3%) of single-arm studies in leading physical therapy journals inappropriately report

treatment effect modifiers, despite using study designs that do not enable this conclusion to be made. This inappropriate reporting risks misleading clinicians and reducing the quality of patient care.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

The authors thank Dr Mark Elkins for providing feedback on a draft version of this manuscript.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.bjpt.2023.100520.

References

1. Qaseem A, Wilt TJ, McLean RM, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the American College of Physicians. *Ann Intern Med.* 2017;166(7):514–530. <https://doi.org/10.7326/M16-2367>.
2. Sun X, Briel M, Busse JW, et al. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ.* 2012;344. <https://doi.org/10.1136/bmj.e1553>.
3. Saragiotto BT, Maher CG, Moseley AM, et al. A systematic review reveals that the credibility of subgroup claims in low back pain trials was low. *J Clin Epidemiol.* 2016;79:3–9. <https://doi.org/10.1016/j.jclinepi.2016.06.003>.
4. Hancock M, Herbert RD, Maher CG. A guide to interpretation of studies investigating subgroups of responders to physical therapy interventions. *Phys Ther.* 2009;89(7):698–704. <https://doi.org/10.2522/ptj.20080351>.
5. Dahan M, Scemama C, Porcher R, et al. Reporting of heterogeneity of treatment effect in cohort studies: a review of the literature. *BMC Med Res Methodol.* 2018;18(1):1–11. <https://doi.org/10.1186/s12874-017-0466-6>.
6. Guccione AA, Goldstein M, Elliott S, et al. Clinical research agenda for physical therapy. *Phys Ther.* 2000;80(5):499–513. <https://doi.org/10.1093/ptj/80.5.499>.
7. Borkan JM, Koes B, Reis S, et al. A report from the second international forum for primary care research on low back pain: reexamining priorities. *Spine.* 1998;23(18):1992–1996. <https://doi.org/10.1097/00007632-199809150-00016>.
8. Assmann SF, Pocock SJ, Enos LE, et al. Subgroup analysis and other (mis) uses of baseline data in clinical trials. *Lancet.* 2000;355(9209):1064–1069. [https://doi.org/10.1016/S0140-6736\(00\)02039-0](https://doi.org/10.1016/S0140-6736(00)02039-0).
9. Gabler NB, Duan N, Liao D, et al. Dealing with heterogeneity of treatment effects: is the literature up to the challenge? *Trials.* 2009;10(1):1–12. <https://doi.org/10.1186/1745-6215-10-43>.
10. Childs JD, Fritz JM, Flynn TW, et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. *Ann Intern Med.* 2004;141(12):920–928. <https://doi.org/10.7326/0003-4819-141-12-200412210-00008>.

11. Kent P, Hancock M. Letter to the Editor concerning "Predicting a beneficial response to motor control training in patients with low back pain: a longitudinal cohort study. by Hides JA, et al. (*Eur Spine J.* 2019; doi. org/10.1007/s00586-019-06045-7) *Eur Spine J.* 2019;28(10):2431. <https://doi.org/10.1007/s00586-019-06122-x>.
12. Hides JA, Murphy M, Jang E, et al. Predicting a beneficial response to motor control training in patients with low back pain: a longitudinal cohort study. *Eur Spine J.* 2019;28(11):2462–2469. <https://doi.org/10.1007/s00586-019-06045-7>.
13. Page M, McKenzie J, Bossuyt P, et al. The PRISMA statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. <https://doi.org/10.1136/bmj.n71>.
14. Sun X, Briel M, Walter SD, et al. Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses. *Bmj.* 2010;340. <https://doi.org/10.1136/bmj.c117>.
15. Hancock MJ, Kent P. Research Note: treatment effect moderators. *J Physiother.* 2022;68(4):283–287.
16. Rothwell PM. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet.* 2005;365(9454):176–186.
17. Hasford J, Bramlage P, Koch G, et al. Inconsistent trial assessments by the National Institute for Health and Clinical Excellence and IQWiG: standards for the performance and interpretation of subgroup analyses are needed. *J Clin Epidemiol.* 2010;63(12):1298–1304. <https://doi.org/10.1016/j.jclinepi.2009.10.009>.