Clinimetrics corner: the many faces of selection bias

Eric J. Hegedus, Jennifer Moody

Duke University, USA

Selection bias, also known as susceptibility bias in an intervention study or spectrum bias in a diagnostic accuracy study, is present throughout clinically applicable evidence in various forms. Selection bias implies that the intervention or diagnostic test has been studied in a less representative sample population, which can lead to inflated overall effect sizes and/or inaccurate findings. Within the literature, there are over 40 forms of selection bias that can influence the external validity of results. Recognition of selection bias is essential in the translation of evidence into effective clinical practice. This clinimetrics corner outlines the major biases that readers encounter and discusses key examples regarding pertinent orthopedic and manual therapy literature.

Keywords: Diagnostic accuracy, External validity, Selection bias, Spectrum effect

Bias is a systematic distortion of the truth. This distortion can be wanton but, more often, is an unintentional side effect of study design, execution, or analysis. In general, the problem with bias is that its presence in a study restricts the generalizability or external validity of that study.¹ A clinician who is a consumer/reader of research may find the performance of a test or intervention, which was represented in a study that involved selection bias, to be different from what he or she experiences in practice. In such a scenario, the clinician may use a reportedly strong intervention with minimal effect or misdiagnose patients based on inflated statistics of diagnostic accuracy of a given test.² The potential consequence is clinical error.

A myriad of biases occur in clinical research. Selection bias, which is also called susceptibility bias in an intervention study³ or spectrum bias in a diagnostic accuracy study,⁴ is one of the more common types of bias encountered in orthopedic manual therapy literature. Spectrum bias can also be referred to as *spectrum effect* because it is so prevalent and influential in studies of diagnostic accuracy.⁴ The terms *susceptibility* and *spectrum bias* imply that the intervention or diagnostic test respectively has been studied in a less representative sample population. The goal of this clinimetrics corner is to further explore these terms, highlight some examples of each in published research studies that are germane to orthopedic manual therapy, and in each section

address how to minimize the occurrence of these biases.

Selection bias and interventions

There are over 40 named forms of selection bias^{3,5} (Tables 1 and 2). The causes of bias involve aggregate categories such as group allocation, disease variability, and sample size deficiencies. What is most compelling is the sheer number of biases involving spectrum bias and the notable challenges in controlling these in a typical study design. The following discussion presents suggestions on how to modify study design to minimize the influence of a few of the biases described in Table 1.

First, a limitation of many studies involving manual therapy is that these studies are often of a single-masked design.^{19,20} Masking or blinding can be performed at the level of the subject, the experimenter, and the data analyst. A single-masked design is commonly one in which either the outcomes assessor or more commonly, the person providing the intervention, is aware of whether the subject is in the control or treatment group. In short, it is impossible to blind the manual therapy practitioner regarding use of correct manipulation versus sham treatment.

Although single-masked designs are a reality in manual therapy,¹⁹⁻²¹ the down side is that this design can inappropriately elevate the intervention effect²²⁻²⁴ and it introduces two more types of bias: ascertainment bias (Table 1) and experimenter or investigator bias. Although subtle, the investigator may believe greatly in the intervention (and not the sham) and unintentionally, convey this expectation of treatment

Correspondence to: Eric Hegedus, Duke University, USA. Email: eric.hegedus@duke.edu

Table 1 Types of selection biases seen primarily in intervention studies

Type of selection bias	Definition
1. Allocation	Non-random assignment to a group based on prognostic variables.
2. Ascertainment	Caused when the investigator responsible for assessing the outcome of interest is also aware of the group assignment. Often an issue in studies based on chart review ⁶ and in single-masked studies.
3. Admission rate	See Berkson's.
4. Authorization	Inability to obtain authorization for release of medical records from certain subjects, which affects the results of medical records research. ⁷
5. Berkson's	In hospital-derived case-control studies, the likelihood of admission may be influenced by the combination of exposure and interest in the disease under study that causes a higher exposure rate among the hospital case subjects than the hospital control subjects. This bias often happens when subjects with more than one condition are more likely to be admitted than patients with a single condition. ⁶
6. Centripetal	A type of healthcare access bias and closely related to referral filter bias. Centripetal bias happens when the reputation of a clinician or clinic attracts a group of subjects not representative of the population as a whole.
7. Competing risks	When a second (faster-acting) disease selectively removes from the population, often by death, persons susceptible to the primary (slower-acting) disease of interest. ⁸
8. Treatment access	A type of healthcare access bias that occurs when subjects have decreased access to healthcare due usually to cultural or socioeconomic reasons.
9. Exclusion	When subjects with a certain exposure (comorbidity) are removed or excluded from either the case subjects or the control group but not both. ⁹
10. Friend control	In case-control studies, can lead to an over-selection of gregarious persons or an overmatching of exposure. ¹⁰
11. Healthcare access	When the subjects admitted to a clinic or institution do not represent the population at large. Types of this bias include popularity, centripetal, diagnostic/treatment access, and referral filter. ⁵
12. Healthy worker effect	There is lower mortality/better health in employed individuals when compared to the general population. ¹¹
13. Incidence-prevalence or Neyman's or selective survival	When the prevalence of disease is underestimated in a study because one of the risk factors influences mortality of subjects.
14. Inclusion	When one or more conditions in control subjects are related to the exposure, producing an inflated exposure in the reference group. ⁵
15. Losses/withdrawals to follow-up	When subjects who withdraw or are lost throughout the study cause a difference in comparison groups. ¹² Withdrawal can be a problem in manual therapy studies that use sham manipulation for the control group.
16. Migrator/migration	When migrants are used as study subjects, they may differ from the population as a whole. Also, different migration habits can introduce bias in epidemiologic studies. ¹³
17. Membership	A group to which the subjects belong (e.g. a running club) might have different health characteristics from the general population.
18. Missing information/data	When missing information about study subjects is ignored or treated as normal or only those records with complete information are selected for inclusion. Non-response to certain questions in surveys and missing values in a database are examples of missing data.
19. Non-contemporaneous control	When a historic control lacks validity because, over time, definitions of disease, diagnostic capabilities, or treatment changes. ¹⁴
20. Non-random sampling	As the title implies, the sampling method is one that produces a sample that is not representative of the population.
21. Non-response	Typically in survey research, potential subjects are not at home, refuse to answer, or are unable to answer. These subjects may be different from those who respond quickly who may be different still from those who respond slowly.
22. Overmatching	Matching of cases and controls is typically done by demographic variables like age, gender, and/or race. These demographics are related to the variable of interest but not of primary interest. However, if the variables are associated with the exposure and not the disease, a selection bias is produced.
23. Post-hoc analysis	Also called a 'fishing expedition' for significance or 'data dredging,' this bias occurs when there is no <i>a priori</i> hypothesis of variable relationship. In a diagnostic accuracy study, a post-hoc choice of a cut point introduces bias. ¹⁵
24. Procedure selection	When treatment assignments are made on the basis of characteristics like co-morbidities or prognostic factors. ¹⁴
25. Purity diagnostic	A type of spectrum bias where subjects with comorbidities are excluded from the study.
26. Referral filter	A type of healthcare access bias, subjects referred to tertiary care facilities like academic medical centers or medical specialists are typically sicker or have rare disorders causing an unusually high concentration of these subjects in studies performed at academic medical centers or by specialists.
27. Relative control	When relatives of the case subjects are enrolled as control subjects, usually in studies of complex diseases with a genetic component. ¹⁶
28. Survivor treatment selection	When survivors of a lethal disease are more likely to enter a study and receive a certain treatment, it is more likely that a positive association will be found between the treatment and survival. ¹⁷
29. Telephone random sampling	A type of non-random sampling bias. Telephone sampling has the benefit of reaching many subjects at relatively low cost, non-telephone households are omitted from the sample, and multi-telephone households have a greater chance of being sampled.
30. Unacceptable disease	Diseases (exposures) that are not acceptable socially are underreported.
31. Volunteer	Sometimes called healthy volunteer bias; in many research designs, those who agree to participate in a study are different from the population in education or health habits like compliance with recommendations, exercise, and the use of preventative medicine.
32. Wrong sample size	Too small a sample size may miss a significant effect when one exists and too large a sample size may establish a statistical significance that is clinically irrelevant. ¹⁴

Table 2 Types of selection biases seen primarily in studies of diagnostic accuracy (some biases are repeated from Table 1)

Type of Selection Bias	Definition
1. Centripetal	A type of healthcare access bias and closely related to referral filter bias. Centripetal bias happens when the reputation of a clinician or clinic attracts a group of subjects not representative of the population as a whole.
2. Detection	When certain variables (exposure) cause an increased/decreased propensity to look for disease rather than diagnose or rule out a disease. Also happens when diagnostic technology improves over time seemingly increasing the incidence of a disease. Types of this bias include diagnostic suspicion, unmasking/detection signal, and mimicry. ⁵
3. Diagnostic access	A type of healthcare access bias. When subjects have decreased access to healthcare due usually to cultural or socioeconomic reasons.
4. Diagnostic suspicion	A type of detection bias that occurs when knowledge of the subject's prior exposure to a putative cause influences the intensity with which the diagnostic process is undertaken.
5. Diagnostic vogue	When a diagnosis is known by more than one label.
6. Healthcare access	When the subjects admitted to a clinic or institution do not represent the population at large. Types of this bias include popularity, centripetal, diagnostic/treatment access, and referral filter. ⁵
7. Mimicry	A type of detection bias that happens when an innocent exposure is seen as causative of the disease of interest simply because it caused a disease similar to the disease of interest. ¹⁴
8. Popularity	A type of healthcare access bias; researchers may track more closely unusual, challenging, or in-vogue diagnoses. ¹⁸
9. Post-hoc analysis	In a diagnostic accuracy study, a post-hoc choice of a cut point introduces bias. ¹⁵
10. Previous opinion	Knowledge of results obtained before examination may alter the examination process and diagnosis in the same patient or a relative. ¹⁴
11. Purity diagnostic	A type of spectrum bias where subjects with comorbidities are excluded from the study.
12. Spectrum	In studies of diagnostic validity, when the subject sample has a limited range of demographics, disease severity, or chronicity, or when true uncertainty about the diagnosis fails to exist as when test performance in subjects with the disease is compared to test performance in subjects who are known not to have the disease.
13. Unmasking or detection signal	A type of detection bias that happens when exposure produces a symptom that improves the likelihood of a correct diagnosis. ¹⁴
14. Wrong sample size	Too small a sample size may miss a significant effect when one exists and too large a sample size may establish a statistical significance that is clinically irrelevant. ¹⁴

efficacy to the subjects, inflating usual self-report outcomes measures such as pain ratings, fear of movement, and function.^{21,25} A study design used to minimize ascertainment and investigator biases was demonstrated by Childs *et al.*²⁶ The study purpose involved validation of a clinical prediction rule (CPR) designed to determine the population that most benefits from lumbar spinal manipulation. A single physical therapist who applied the CPR was masked to group assignment; the examiners were not aware of the CPR's criteria or the subject's status based on the CPR, and the outcomes assessor was likewise masked.

The Childs *et al.*'s²⁶ study, though well designed, also provides an example of two types of selection bias called membership bias and referral filter bias (Table 1). Both biases involve the selection of patients from a location who are potentially dissimilar to those in a general population from other, more common locations. The subjects for the Childs *et al.*'s²⁶ study were primarily recruited from a number of United States Air Force facilities and two academic medical centers. One could argue that subjects in the United States Air Force have a health profile that differs from the population in general. This argument has been supported by a follow-up study.²⁷ Hancock and colleagues²⁷ examined the Childs *et al.*'s²⁶ CPR in a general population presenting to primary care practices and found that the CPR performed no better than chance. Although there were differences in the interventional methods

of both studies, the profound difference in results demonstrates the potential effect of membership and referral filter bias.

A more obvious example of referral filter bias in manual therapy literature is detailed in an article by Haldeman *et al.*²⁸ These authors investigated the difference in perception between chiropractors and neurologists regarding the risk of vertebral artery dissection after manipulation. One of every two neurologists was aware of a case of vertebral artery dissection post-manipulation whereas one of every 48 chiropractors was aware of vascular incident or stroke in their practice lifetime. The Haldeman *et al.*'s²⁸ study required some extrapolation but the dramatic difference in perception represents and is due to referral filter bias since the patient with a vascular incident after manipulation will see three MD specialists in neurology for every one chiropractor.

In addition to membership and referral filter bias, wrong sample size bias can be a limitation in manual therapy studies. As an example, a systematic review on the effectiveness of spinal mobilization and manipulation in patients with headache showed that sample size was too small in all eight studies.¹⁹ Insufficient sample size can result in a misrepresentation of the effectiveness of a tested intervention either lessening the effect or finding no statistical significance when one might truly exist with a larger sample size.

While membership bias, referral filter bias, and wrong sample size are influential selection biases of

which the reader should be aware, a more common selection bias in manual therapy literature is non-random sampling. For example, in a recent systematic review of manual therapy interventions in patients with lumbar spinal stenosis,²⁹ only 1 of 11 studies made use of a randomized clinical design. Worth repeating is that although the effect of non-randomization can vary, study designs that do not randomize the assignment of subjects to different treatments are generally found to overestimate the effect of the intervention.²⁴ The importance of randomization cannot be overstated as it is the single greatest protection against selection bias in all of its subtle forms.²⁴

Selection bias and diagnostic accuracy

Unfortunately, diagnosis is not an area of study that readily lends itself to the randomized clinical trial since subjects are usually chosen based on a specific set of signs or symptoms that lead the clinician to suspect the presence of a certain type of pathology. The effect of spectrum bias is so strong in studies of diagnostic accuracy that it is now widely accepted that estimates of test accuracy like sensitivity, specificity, and likelihood ratios will be dependent upon the population in which the test or measure is applied.^{4,30,31} Based on this realization, the recommendation to use the term spectrum effect has emerged, indicating that a variation in test performance is very relevant and a finding that is almost certainly to affect the outcome of the study.

In our experience, a group of tests that is highly prone to spectrum bias/effect is orthopedic special tests of the shoulder, because women are greatly under-represented in these studies in general.³² For example, the pain provocation test was introduced by Mimori *et al.*³³ to detect a superior labral anterior-to-posterior lesion. The test was much heralded with a reported sensitivity of 100% and a specificity of 90%. The spectrum of patients, however, involved 32 overhead athletes at the ages of 17–29, a group that included two females and 30 males. In a subsequent study, Parentis *et al.*³⁴ recruited a broader group of patients (ages 15–71) with a better male-to-female ratio and the reported sensitivity of the pain provocation test dropped to 17%. Both of these studies were of limited quality so the change in diagnostic statistics cannot be attributed to spectrum bias alone, but the research on this test is just one of many examples of the changing values of sensitivity and specificity due to the spectrum effect.

The argument for not correcting for spectrum bias but simply accepting the spectrum effect has been made.³⁰ However, the Standards for Reporting of Diagnostic Accuracy Group² and Whiting *et al.*³⁵ have recommended a prospective, consecutive subjects design with extensive description of the subjects and the method of their recruitment to minimize the

effect of spectrum bias. Furthermore, a key component to minimizing spectrum bias is to examine the use of any test in the same manner as that test would be used in the clinic, i.e. in the context of true diagnostic uncertainty. Using a control group that is either healthy or has a completely unrelated diagnosis is the greatest source of overestimation of the diagnostic performance of a test.^{4,36} Simply stated, if a clinician is interested in discovering the usefulness of a new test for a torn rotator cuff, a control group of patients on the waiting list for a condition disparate to that of a shoulder problem (e.g. a patient waiting for lumbar surgery) is a poor control group.

Discussion

Whether in a study of diagnostic accuracy or a study of intervention, spectrum bias and the threat to external validity that this bias produces create error when the consumer of published literature attempts to use the findings in his or her practice setting without recognition of potential biases. The forms of selection bias are so numerous and subtle that it is unlikely that a clinician can be made aware of all forms. Perhaps the best hope is awareness of a research design that guards against selection bias such as use of prospective recruitment, randomization, and masking.

In studies involving diagnostic accuracy, we may simply have to accept spectrum bias as a normal occurrence, an effect to be recognized. Variations of subject population will alter test findings just as comparative groups (such as those used in case control designs) can inflate test accuracies. In either case, an awareness of the selection of study subjects and their controls is vital to orthopedic manual therapists who strive to be on the leading edge of evidence-based practice as they apply research findings to everyday practice.

Conclusion

Selection bias is a common form of bias in both interventional and diagnostic accuracy studies. Controlling for bias involves use of masking, random design, proper case control groups in situations of diagnostic uncertainty, and due diligence in controlling (and reporting) biases in all studies. Because all biases cannot be controlled, it is imperative that clinicians understand the selection biases that most commonly influence study outcomes.

References

- 1 Stang A. Appropriate epidemiologic methods as a prerequisite for valid study results. *Eur J Epidemiol.* 2008;23:761–5.
- 2 Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al.* The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Clin Chem.* 2003;49:7–18.
- 3 Hartman JM, Forsen JW, Jr., Wallace MS, Neely JG. Tutorials in clinical research: Part IV: recognizing and controlling bias. *Laryngoscope.* 2002;112:23–31.

- 4 Lijmer JG, Mol BW, Heisterkamp S, *et al*. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA*. 1999;282:1061–6.
- 5 Delgado-Rodriguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004;58:635–41.
- 6 Sutton-Tyrrell K. Assessing bias in case-control studies: proper selection of cases and controls. *Stroke*. 1991;22:938–42.
- 7 Jacobsen SJ, Xia Z, Campion ME, Bonsel GJ, Prins MH, van der Meulen JH, *et al*. Potential effect of authorization bias on medical record research. *Mayo Clin Proc*. 1999;74:330–8.
- 8 Schatzkin A, Slud E. Competing risks bias arising from an omitted risk factor. *Am J Epidemiol*. 1989;129:850–6.
- 9 Horwitz RI, Feinstein AR. Exclusion bias and the false relationship of reserpine and breast cancer. *Arch Intern Med*. 1985;145:1873–5.
- 10 Ross JA, Spector LG, Olshan AF, Bunin GR. Invited commentary: birth certificates—A best control scenario? *Am J Epidemiol*. 2004;159:922–4; discussion 925.
- 11 Arrighi HM, Hertz-Picciotto I. The evolving concept of the healthy worker survivor effect. *Epidemiology*. 1994;5:189–96.
- 12 Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol*. 1977;106:184–7.
- 13 Tong S. Migration bias in ecologic studies. *Eur J Epidemiol*. 2000;16:365–9.
- 14 Sackett DL. Bias in analytic research. *J Chronic Dis*. 1979;32:51–63.
- 15 Leeflang MM, Moons KG, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: mechanisms, magnitude, and solutions. *Clin Chem*. 2008;54:729–37.
- 16 Goldstein AM, Hodge SE, Haile RW. Selection bias in case-control studies using relatives as the controls. *Int J Epidemiol*. 1989;18:985–9.
- 17 Sy R, Bannon P, Bayfield M, Brown C, Kritharides L. Survivor treatment selection bias and outcomes research. *Circ Cardiovasc Qual Outcomes*. 2009;2:469–74.
- 18 Kelly S, Berry E, Roderick P, Harris KM, Cullingworth J, Gathercole L, *et al*. The identification of bias in studies of the diagnostic performance of imaging modalities. *Br J Radiol*. 1997;70:1028–35.
- 19 Fernandez-de-las-Peñas C, Alonso-Blanco C, San-Roman J, Miangolarra-Page JC. Methodological quality of randomized controlled trials of spinal manipulation and mobilization in tension-type headache, migraine, and cervicogenic headache. *J Orthop Sports Phys Ther*. 2006;36:160–9.
- 20 Furlan AD, Imamura M, Dryden T, Irvin E. Massage for low back pain: an updated systematic review within the framework of the Cochrane Back Review Group. *Spine*. 2009;34:1669–84.
- 21 Licciardone JC, Russo DP. Blinding protocols, treatment credibility, and expectancy: methodologic issues in clinical trials of osteopathic manipulative treatment. *J Am Osteopath Assoc*. 2006;106:457–63.
- 22 Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. *Ann Intern Med*. 2001;135:982–9.
- 23 Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ*. 1998;317:1185–90.
- 24 Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev*. 2007;(12):MR000012.
- 25 Houben R, Ostelo R, Vlaeyen J, Wolters P, Peters M, Stomp-van den Berg S. Health care providers' orientations toward common low back pain predict perceived harmfulness of physical activities and recommendations regarding return to normal activity. *Eur J Pain*. 2005;9:173–83.
- 26 Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR, *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;141:920–8.
- 27 Hancock MJ, Maher CG, Latimer J, Herbert RD, McAuley JH. Independent evaluation of a clinical prediction rule for spinal manipulative therapy: a randomised controlled trial. *Eur Spine J*. 2008;17:936–43.
- 28 Haldeman S, Carey P, Townsend M, Papadopoulos C. Clinical perceptions of the risk of vertebral artery dissection after cervical manipulation: the effect of referral bias. *Spine*. 2002;23:334–42.
- 29 Reiman MP, Harris JY, Cleland J. Manual therapy interventions for patients with lumbar spinal stenosis: a systematic review. *NZ J Physio*. 2009;37:17–28.
- 30 Mulherin SA, Miller WC. Spectrum bias or spectrum effect? Subgroup variation in diagnostic test evaluation. *Ann Intern Med*. 2002;137:598–602.
- 31 Goehring C, Perrier A, Morabia A. Spectrum bias: a quantitative and graphical analysis of the variability of medical diagnostic test performance. *Stat Med*. 2004;23:125–35.
- 32 Hegedus EJ, Goode A, Campbell S, Morin A, Tamaddoni M, Moorman CT, *et al*. Physical examination tests of the shoulder: a systematic review with meta-analysis of individual tests. *Br J Sports Med*. 2008;42:80–92; discussion 92.
- 33 Mimori K, Muneta T, Nakagawa T, Shinomiya K. A new pain provocation test for superior labral tears of the shoulder. *Am J Sports Med*. 1999;27:137–42.
- 34 Parentis MA, Glousman RE, Mohr KS, Yocum LA. An evaluation of the provocative tests for superior labral anterior posterior lesions. *Am J Sports Med*. 2006;34:265–8.
- 35 Whiting P, Rutjes AW, Dinnes J, Reitsma J, Bossuyt PM, Kleijnen J. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess*. 2004;8:iii, 1–234.
- 36 Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ*. 2006;174:469–76.