Curr Anesthesiol Rep (2016) 6:293–298
DOI 10.1007/s40140-016-0174-5

Causal Inference in Anesthesia and Perioperative Observational Studies

Tri-Long Nguyen[1,2] · Audrey Winter[1] · Jessica Spence[3] · Géraldine Leguelinel-Blache[1,2] · Paul Landais[1,4] · Yannick Le Manach[3,5]

Published online: 9 July 2016

Abstract
Purpose of Review Observational studies are of great importance to anesthesia and perioperative care research, as they reflect routine clinical practice. However, because observational data are nonexperimental, assigning causality to identified relationships has a significant risk of bias. After describing the pros and cons of observational studies, we provide an overview of the different methods used to make causal inferences. Of these, we focus on the propensity score analysis, which achieves an increasing popularity in anesthesia and perioperative literature.
Recent Findings Several methods are proposed for estimating treatment effects in observational studies. Although multivariable regression has traditionally been used to infer causal effects by adjusting for confounding variables, the reported result mainly depends on the model specification that fits the researcher's hypothesis. Preprocessing observational data can reduce this model dependence, by balancing confounders across the treatment groups like in experimental studies. In particular, the propensity score analysis approximates the randomized controlled trial.
Summary Compared to randomized experiments, observational studies are low-cost sources of "real-life" data, but they are exposed to bias. Treatment effects can be estimated by using appropriate methods, such as the propensity score analysis, which limits confounding and model-dependence bias. We provide an illustrative example of propensity score analysis using a recently published study, which assessed the outcomes after hip fracture surgery compared with elective total hip replacement.

Keywords Causal inference · Observational study · Propensity score · Anesthesia · Perioperative care

This article is part of the Topical collection on *Research Methods and Statistical Analyses*.

✉ Tri-Long Nguyen
longbacon.nguyen@gmail.com

[1] Laboratory of Biostatistics, Epidemiology, Clinical Research and Health Economics, EA2415, Faculty of Medicine, University of Montpellier, Montpellier, France

[2] Department of Pharmacy, Nîmes University Hospital, CHRU Nîmes, Service de Pharmacie, 4 Rue du Professeur Robert Debré, 30029 Nîmes, France

[3] Departments of Anesthesia, Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, Michael DeGroote School of Medicine, McMaster University, Hamilton, ON, Canada

[4] Department of Biostatistics, Epidemiology, Public Health and Medical Informatics, Nîmes University Hospital, Nîmes, France

[5] Perioperative Research Group, Population Health Research Institute, Hamilton, ON, Canada

Introduction

The randomized controlled trial (RCT) is often perceived as the standard method to provide 'evidence' of a treatment effects [1]. However, there are many situations in which RCT evaluation is not feasible, as in conditions requiring emergency management or where random allocation of interventions is unethical or impossible [2]. In these situations, observational studies should be considered. Like RCTs, observational studies aim to estimate the causal effect of a treatment or an intervention. Unlike RCTs, observational studies collect "real-life" data, and have the potential to provide results that are reflective of routine clinical practice. Compared to RCTs, observational studies are low-cost sources of data and can enroll numerous

patients [3]. However, because treatment allocation is not randomized, the estimation of causal effects has a significant risk of bias. After describing these biases, we present an overview of the different methods to minimize their influence on treatment effect estimates.

Bias Due to Nonrandomization

As opposed to randomized experiments, the treatment assignment in observational studies is determined by a clinical decision rather than by random allocation. Subsequently, the population selected to receive the treatment may systematically differ from the control population [4]. These differences may lead to two types of bias: confounding and selection bias.

As an illustrative example, we take a recent study published by Le Manach et al. assessing the effect of hip fracture surgery compared to total hip replacement on postoperative in-hospital mortality [5••]. In this situation, randomization was unethical, as previous studies have suggested an association between hip fracture surgery and increasing risk of mortality [6]. In this observational study of 690,995 patients [5], hip fracture surgery was performed preferentially on older patients with more preoperative comorbidities. A crude estimation revealed a risk ratio (RR) equal to 18.96 (95 % CI 17.54–20.50, $P < 0.001$). However, such an analysis confounds the effect of hip fracture surgery with those of age and preoperative comorbidities. After matching on preoperative variables (age, sex, comorbidities, residency, and center type), the authors reported a RR = 5.88 (95 % CI 5.26–6.58, $P < 0.001$), which reflects the specific effect of hip fracture surgery measured on the older, multi-morbid patients who underwent this procedure. Because of the fundamental differences in the hip fracture surgery group, the effect noted may apply only to this population and may not generalize to the population as a whole.

On the one hand, confounding bias distorts the apparent treatment effect either quantitatively (i.e., overestimation or underestimation of the treatment effect) or qualitatively (i.e., apparent benefit of a truly harmful treatment). On the other hand, selection bias leads one to wrongly extrapolate a conclusion drawn from a specific population (e.g., the treated population) to a different population (e.g., the overall population). In large RCTs, random allocation guarantees the removal of systematic differences between treatment and control groups [4]. Because groups are assumed to be equivalent regarding all prognostic factors, no confounding bias is expected. Because the treated population is assumed to be equal to the control population, and therefore to the overall population, no selection bias is expected. In RCTs, the average treatment effect in the overall population (ATE) is thus equal to the average treatment effect in the treated population (ATT) and to the average treatment effect in the control population (ATC). In 1997, Heckman defined the selection bias, as the difference between the ATE and the ATT [7]. In observational studies, the ATE, ATT, and ATC can be estimated, but are not equal. To avoid selection bias, the choice of the estimand should be guided by the clinical objective of the study. The ATT is often referred to as the estimand of interest for decision-making, because it is measured on subjects for whom the intervention was intended [7]. To address confounding bias in observational studies, several methods of treatment effect estimation are described in the literature. Of these, we note the importance of propensity score analysis, which has been increasingly used in anesthesia and perioperative research [8].

Methods to Estimate Causal Effects

'Traditional' Regression Adjustment

Multivariable regression has traditionally been used to infer causal effects by adjusting for confounding variables (e.g., age, sex, comorbidities, residency, and center type). Because it involves statistical modeling, this approach requires adherence to the assumptions of the statistical model used. Logistic and survival models (which are used for binary and time-to-event outcomes, respectively) should include a minimum of 10 events per variable for reliable estimation [9–11], though it may be possible to obtain accurate estimates with slightly fewer [12, 13]. When linear regression is used to evaluate a continuous outcome, the need for a minimum of two subjects per variable has been described [14]. 'Traditional' regression adjustment is therefore limited in small samples, or when the outcome of interest is rare. In these situations, a sound strategy of variable selection becomes crucial. Several popular approaches to automatic variable selection are described, though these have been found to produce unstable results and require validation using resampling methods [15]. Because effect estimates are obtained from statistical models, the regression adjustment assumes their accuracy and goodness-of-fit. An alternative to this 'traditional' approach is the propensity score analysis, which aims to create a balanced distribution of confounders, similar to what one would expect in an RCT [16].

Propensity Score Analysis

The propensity score is defined as the conditional probability of receiving the treatment of interest (e.g., hip fracture surgery) given a set of confounding covariates (e.g., age, sex, comorbidities, residency, and center type) [16].

This single score summarizes the values of all confounders, meaning that subjects with equal scores are expected to have, on average, similar values for each confounding variable [16]. For example, patients with equal propensity to undergo hip fracture surgery are expected to be of similar age, sex, center type, and residency, and to have the same comorbidities. Therefore, treated and control subjects with similar propensity scores are balanced regarding confounders and differ only regarding treatment status. This corresponds to a counterfactual framework, sometimes called "quasi-randomization." Like in RCTs, wherein all subjects have a propensity score of 0.5 (assuming equal and random allocation), such a design allows for an unbiased estimation of treatment effect. Propensity score analysis comprises 4 steps:

1) Estimation of propensity scores;
2) Creating a balanced framework that compares treated and control patients with similar propensity scores;
3) Assessing the balance of covariates across treated and control groups; and
4) Estimation of the treatment effect.

First, assuming that all confounders have been measured, the propensity score of each subject can be estimated through regression methods. Typically, logistic regression is performed, with the treatment status designated as the outcome of interest and the confounders included as explanatory variables. It is recommended that only true confounders (i.e., related to both the treatment allocation and the outcome) or prognostic covariates (i.e., related to the outcome) are included in the propensity score model [17], as the inclusion of other covariates may result inflation of bias [18]. Because the aim of propensity score analysis is not predictive, but rather to balance group allocation, the discriminative performance of the model should not guide model building [19]. However, in spite of these recommendations, strategies for covariate selection in propensity score-based studies remain poorly described in the literature [20••]. In addition to logistic regression, there is a wide variety of modeling approaches, which include data-mining [21], machine learning [22] or covariates-balancing [23, 24].

Several methods are available to compare subjects with similar propensity scores across treatment groups, including adjustment, matching, stratification, and weighting [25]. Of these, matching and weighting have been demonstrated to perform better, in terms of bias and variance minimization [26•, 27]. In weighting, pseudo-populations are created using different weights for treated and controls, according to the estimand of interest. To estimate the ATE, inverse probability of treatment weighting (IPTW) is used as follows: $W = \frac{1}{PS}$ and $W = \frac{1}{1-PS}$ for treated and controls, respectively [25]. To estimate the ATT: $W = 1$ and $W = \frac{PS}{1-PS}$, and to estimate the ATC: $W = \frac{1-PS}{PS}$ and $W = 1$, respectively [34].

In matching, pairs of treated and control patients with similar propensity scores are created. Propensity scores of each member must be within a prespecified distance of one another, which is referred as the caliper width [28, 29]. Subjects for whom no match are found are discarded from the matched sample, leading to an incomplete matching analysis. Therefore, the caliper width must be set at a value that minimizes both the bias due to the distance in propensity score between treated and controls [28, 29], and that due to incomplete matching [30]. A maximal caliper width equal to 0.2 of the standard deviation of the logit of the propensity score has been described as best to ensure balance [28], though narrower width may be preferable [5, 29•]. To estimate the ATT, the matching ratio (i.e., the number of controls per treated subjects) can vary from 1:1 to 1:M, with or without replacement (i.e., with or without reusing controls, which have been previously matched). Without replacement, a ratio of 1:1 performs better than 1:M [31]. Several matching algorithms have been described, including nearest neighbor or "greedy" matching and optimal matching, though no substantial differences have been found in terms of their performance [32•]. Other methods, such as matching on Mahalanobis distance, are less commonly used in the medical literature [8, 20, 33]. In our example, Le Manach et al. conducted a 1:1 greedy matching on propensity scores, without replacement, using a caliper width of 10^{-5} to create a counterfactual framework [5].

Once the treatment and control groups are established, balance of covariates across the matched (or weighted) sample must be evaluated. Because balance is a property of the sample and not the population, balance metrics should be reflective of the sample alone and should not be influenced by sample size [35]. As in RCTs, there is therefore no reason to conduct significance tests to assess the differences in covariates across treatment groups [36–40]. Standardized mean differences are more commonly used for balance diagnostics, and can be defined as

$$SMD = 100 \times \frac{|\bar{x}_{treated} - \bar{x}_{control}|}{\sqrt{\frac{s^2_{treated} + s^2_{control}}{2}}}$$

$\bar{x}$ denoting the variable mean (or proportion for binary variables and classes of categorical variables) and s^2 is the variance. While some authors define imbalance as an SMD > 10 % [41], others consider a threshold cut-off of 5 % [5]. Although there is no consensus as to the value of the cut-off chosen, it is recommended to minimize differences between groups as much as possible [35].

Regardless, balance checking in propensity score analysis is of central importance, and, according to a recent systematic review, its reporting is often suboptimal in medical studies [20]. In our study case, Le Manach et al. reported standardized mean differences below 1 % for all covariates after matching [5].

Once the balance of covariates between matched (or weighted) samples is ensured, treated and controls can be compared without confounding, and the treatment effect can be estimated by taking account the paired nature of the data [42]. Like in RCTs, any remaining unbalanced confounders can be included as covariates along with treatment status in a second regression (sometimes called "double-robust" approach). In their study, Le Manach et al. estimated a harmful effect of hip fracture surgery on mortality using a conditional Poisson regression on the matched sample: RR = 5.88 (95 % CI 5.26–6.58) [5].

Propensity Score Analysis or 'Traditional' Adjustment?

In contrast to the 'traditional' approach using regression to adjust for confounders, propensity score analysis creates a new, quasi-randomized framework. For this reason, propensity score analysis is considered valid whether or not balance of covariates is achieved. However, there are certain situations in which the 'traditional' and propensity score approaches are not equally applicable. When evaluating binary or time-to-event outcomes, logistic and survival models provide conditional effects, while propensity score analysis estimates marginal effects as one would in an RCT. These two effects differed in definition. The conditional effect corresponds to the mean difference (or ratio) in the value of the outcome for all the subjects, whether they were treated and not, whereas the marginal effect corresponds to the difference (or ratio) in mean value of the outcome between each treated and untreated subject. These two effects coincide on linearized scale (i.e., absolute risk difference), but not on odds or hazard ratio scales [43]. There are therefore fundamental differences in estimates of treatment effect when comparing logistic regression and propensity score analysis [44]. Estimating marginal effects using logistic and survival models is possible [45, 46], although the use of 'traditional' regression may have the potential for researcher's bias [47].

Other Methods

Other approaches that are important to mention include disease-risk scores, which are multivariate models predicting the occurrence of an outcomes, conditional on nonexposure to the treatment [48]. Similar to a propensity score, a disease-risk score summarizes a set of confounding covariates into a single score. Because treated and control individuals of the same baseline risk are comparable, regardless of treatment status, adjusted, stratified or matched analyses using disease-risk scores can be therefore conducted to remove confounding bias [49, 50]. Currently, however, there is limited interest in the medical research community in the use of these tools [51].

Conclusion

Observational studies, because nonexperimental, are exposed to selection and confounding bias. This article provides an overview of the methods which can be used to estimate unbiased treatment effects. In particular, we focused on propensity score analysis, a method that is becoming increasingly popular in anesthesia and perioperative care research. In contrast to 'traditional' regression adjustment, this approach aims to create balanced frameworks in observational studies, so as to estimate unbiased causal effects similarly to randomized experiments.

Compliance with Ethical Guidelines

Conflict of Interest Tri-Long Nguyen, Audrey Winter, Jessica Spence, Géraldine Leguelinel-Blache, Paul Landais, and Yannick Le Manach declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent This article does not contain any studies with human or animal subjects performed by any of the authors.

References

Papers of particular interest, published recently, have been highlighted as:
- • Of importance
- •• Of major importance

1. Guyatt GH, Oxman AD, Kunz R, Vist GE, Falck-Ytter Y, Schunemann HJ, Group GW. What is "quality of evidence" and why is it important to clinicians? BMJ. 2008;336(7651):995–8.
2. Naylor CD, Guyatt GH. Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services. The evidence-based medicine working group. JAMA. 1996;275(7):554–8.
3. Feinstein AR. Epidemiologic analyses of causation: the unlearned scientific lessons of randomized trials. J Clin Epidemiol. 1989;42(6):481–489; discussion 499–502.
4. Altman DG, Bland JM. Statistics notes. Treatment allocation in controlled trials: why randomise? BMJ. 1999;318(7192):1209.
5. •• Le Manach Y, Collins G, Bhandari M, Bessissow A, Boddaert J, Khiami F, Chaudhry H, De Beer J, Riou B, Landais P et al. Outcomes after hip fracture surgery compared with elective total hip replacement. JAMA. 2015;314(11):1159–66. *In a large cohort of French patients, hip fracture surgery compared with elective total hip replacement is associated with a higher risk of*

in-hospital mortality after matching on age, sex and measured comorbidities.
6. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation Study I, Devereaux PJ, Chan MT, Alonso-Coello P, Walsh M, Berwanger O, Villar JC, Wang CY, Garutti RI, Jacka MJ et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. JAMA. 2012;307(21):2295–304.
7. Heckman J. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. J Hum Resour. 1997;32(3):441–62.
8. Gayat E, Pirracchio R, Resche-Rigon M, Mebazaa A, Mary JY, Porcher R. Propensity scores in intensive care and anaesthesiology literature: a systematic review. Intensive Care Med. 2010;36(12):1993–2003.
9. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49(12):1373–9.
10. Concato J, Peduzzi P, Holford TR, Feinstein AR. Importance of events per independent variable in proportional hazards analysis. I. Background, goals, and general strategy. J Clin Epidemiol. 1995;48(12):1495–501.
11. Peduzzi P, Concato J, Feinstein AR, Holford TR. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. J Clin Epidemiol. 1995;48(12):1503–10.
12. Cepeda MS. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. Am J Epidemiol. 2003;158(3):280–7.
13. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. Am J Epidemiol. 2007;165(6):710–8.
14. Austin PC, Steyerberg EW. The number of subjects per variable required in linear regression analyses. J Clin Epidemiol. 2015;68(6):627–36.
15. Austin PC, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. J Clin Epidemiol. 2004;57(11):1138–46.
16. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41–55.
17. Austin PC, Grootendorst P, Anderson GM. A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. Stat Med. 2007;26(4):734–53.
18. Pearl J. Invited commentary: understanding bias amplification. Am J Epidemiol. 2011;174(11):1223–7; discussion 1228–9.
19. Westreich D, Cole SR, Funk MJ, Brookhart MA, Sturmer T. The role of the c-statistic in variable selection for propensity score models. Pharmacoepidemiol Drug Saf. 2011;20(3):317–20.
20. •• Ali MS, Groenwold RH, Belitser SV, Pestman WR, Hoes AW, Roes KC, Boer A, Klungel OH. Reporting of covariate selection and balance assessment in propensity score analysis is suboptimal: a systematic review. J Clin Epidemiol. 2015;68(2):112–21. *The execution and reporting of propensity score analysis is far from optimal in medical literature.*
21. Setoguchi S, Schneeweiss S, Brookhart MA, Glynn RJ, Cook EF. Evaluating uses of data mining techniques in propensity score estimation: a simulation study. Pharmacoepidemiol Drug Saf. 2008;17(6):546–55.
22. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. Stat Med. 2010;29(3):337–46.
23. Imai K, Ratkovic M. Covariate balancing propensity score. J R Stat Soc B. 2014;76(1):243–63.
24. Wyss R, Ellis AR, Brookhart MA, Girman CJ, Jonsson Funk M, LoCasale R, Sturmer T. The role of prediction modeling in propensity score estimation: an evaluation of logistic regression, bCART, and the covariate-balancing propensity score. Am J Epidemiol. 2014;180(6):645–55.
25. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat Med. 2004;23(19):2937–60.
26. • Austin PC. The performance of different propensity score methods for estimating marginal hazard ratios. *Statistics in medicine* 2013;32(16):2837–49. *Matching and inverse probability of treatment weighting methods demonstrate better performances for estimating marginal hazard ratios.*
27. Austin PC. The performance of different propensity-score methods for estimating differences in proportions (risk differences or absolute risk reductions) in observational studies. Stat Med. 2010;29(20):2137–48.
28. Austin PC. Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharm Stat. 2011;10(2):150–61.
29. • Lunt M: Selecting an appropriate caliper can be essential for achieving good balance with propensity score matching. Am J Epidemiol. 2014;179(2):226–35. *A tigher caliper width leads to reduced bias and closer matches in propensity score matching analysis.*
30. Rosenbaum PR, Rubin DB. The bias due to incomplete matching. Biometrics. 1985;41(1):103–16.
31. Austin PC. Statistical criteria for selecting the optimal number of untreated subjects matched to each treated subject when using many-to-one matching on the propensity score. Am J Epidemiol. 2010;172(9):1092–7.
32. • Austin PC. A comparison of 12 algorithms for matching on the propensity score. Stat Med. 2014;33(6):1057–69. *Nearest neighbor matching induces the same balance in baseline covariates as does optimal matching. Caliper matching tends to result in estimates of treatment effect with less bias compared with optimal and nearest neighbor matching.*
33. Austin PC. Propensity-score matching in the cardiovascular surgery literature from 2004 to 2006: a systematic review and suggestions for improvement. J Thorac Cardiovasc Surg. 2007;134(5):1128–35.
34. Morgan SL, Todd JL. A diagnostic routine for the detection of consequential heterogeneity of causal effects. Sociol Methodol. 2008;38:231–81.
35. Imai K, King G, Stuart EA. Misunderstandings among experimentalists and observationalists about causal inference. J R Stat Soc Ser A (Statistics in Society). 2008;171(2):481–502.
36. Altman DG, Dore CJ. Randomisation and baseline comparisons in clinical trials. Lancet. 1990;335(8682):149–53.
37. Assmann SF, Pocock SJ, Enos LE, Kasten LE. Subgroup analysis and other (mis)uses of baseline data in clinical trials. Lancet. 2000;355(9209):1064–9.
38. Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Stat Med. 2002;21(19):2917–30.
39. Senn S. Testing for baseline balance in clinical trials. Stat Med. 1994;13(17):1715–26.
40. Senn SJ. Covariate imbalance and random allocation in clinical trials. Stat Med. 1989;8(4):467–75.
41. Normand ST, Landrum MB, Guadagnoli E, Ayanian JZ, Ryan TJ, Cleary PD, McNeil BJ. Validating recommendations for coronary angiography following acute myocardial infarction in the elderly: a matched analysis using propensity scores. J Clin Epidemiol. 2001;54(4):387–98.

42. Austin PC. Comparing paired vs non-paired statistical methods of analyses when making inferences about absolute risk reductions in propensity-score matched samples. Stat Med. 2011;30(11):1292–301.
43. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. Am J Epidemiol. 1987;125(5):761–8.
44. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Systematic differences in treatment effect estimates between propensity score methods and logistic regression. Int J Epidemiol. 2008;37(5):1142–7.
45. Austin PC. Absolute risk reductions, relative risks, relative risk reductions, and numbers needed to treat can be obtained from a logistic regression model. J Clin Epidemiol. 2010;63(1):2–6.
46. Austin PC. Absolute risk reductions and numbers needed to treat can be obtained from adjusted survival models for time-to-event outcomes. J Clin Epidemiol. 2010;63(1):46–55.
47. Rubin DB. Using propensity scores to help design observational studies: application to the tobacco litigation. Health Serv Outcomes Res Method. 2001;2:169–88.
48. Miettinen OS. Stratification by a multivariate confounder score. Am J Epidemiol. 1976;104(6):609–20.
49. Sturmer T, Schneeweiss S, Brookhart MA, Rothman KJ, Avorn J, Glynn RJ. Analytic strategies to adjust confounding using exposure propensity scores and disease risk scores: nonsteroidal antiinflammatory drugs and short-term mortality in the elderly. Am J Epidemiol. 2005;161(9):891–8.
50. Wyss R, Ellis AR, Brookhart MA, Jonsson Funk M, Girman CJ, Simpson RJ, Jr., Sturmer T. Matching on the disease risk score in comparative effectiveness research of new treatments. Pharmacoepidemiol Drug Saf. 2015;24(9):951–61.
51. Tadrous M, Gagne JJ, Sturmer T, Cadarette SM. Disease risk score as a confounder summary method: systematic review and recommendations. Pharmacoepidemiol Drug Saf. 2013;22(2):122–9.