Forschende Komplementärmedizin und Klassische Naturheilkunde

Forsch Komplementärmed Klass Naturheilkd 2000;11(suppl 1):30–35

Treatment Outcomes in Patients with Migraine: An Ex-Post-Facto Comparison of Two In-Patient Facilities

W. Weidenhammer[a] K. Linde[a] D. Melchart[a,b]

[a] Zentrum für naturheilkundliche Forschung, II. Med. Klinik und Poliklinik, TU München, Germany
[b] Abteilung Naturheilkunde, Department für Innere Medizin, Universitätsspital Zürich, Switzerland

Key Words

Non-randomized comparative study · Case-mix adjustment techniques · Propensity score · Covariance analysis · Migraine

Summary

Background: In two hospitals we performed an open, prospective observational study on patients with chronic headache as a measure of internal quality assurance using identical methods. Available data were subordinately analysed in order to compare both studies. **Questions:** Are the patient samples of both hospitals comparable? If not, which form of statistical adjustment is recommended? Are there differences in the outcome measures of both facilities? **Methods:** The outcomes were defined as differences between baseline values and values at discharge from hospital, respectively 6 months after. Frequency of headache attacks, intensity of pain, intensity of general complaints as well as of concomitant symptoms, and quality of life were determined in advance as dependent variables. To compare both patient groups univariate analysis of variance without and with inclusion of covariates were used. For calculating propensity scores (conditional probability of belonging to one of two groups) a logistic regression with the same covariates serving as independent variables was performed. **Patients:** 426 patients with the main diagnosis 'Migraine' and complete data sets concerning the covariates were selected for analysis. 87% of patients are female, the mean age is 45.5 ± 11.7 years (range 14–73 yrs). **Results:** 4 out of 11 potential covariates show statistically significant differences between the patients of both hospitals. Correct classification of patients by means of the propensity score succeeds in 67%. Comparing the outcomes at discharge from hospital, significant differences between both groups exist which are, with one exception, not affected by controlling for covariates. 6 months after discharge two of the initial differences between both patient groups are no longer present. These findings are independent from the statistical technique of risk adjustment. **Conclusions:** Because of the observed differences between both patient groups it is recommended to adjust data by regression analysis in order to enhance comparability. The choice for one of the two proposed techniques is secondary. With respect to the analyses clear differences between both hospitals exist in short-term outcomes, disappearing 6 months later.

Rationale

The German Social Security Code requires hospitals to implement quality assurance measures to enable the conduct of comparative studies (Section 137, Social Security Code V). Considering that it is rarely possible to conduct studies with randomised assignment of patients referred to different care-providers (hospitals), the question of the adequate methodological procedure is increasingly important. Besides the general questions concerning the quality criteria for non-randomised studies, or even for so-called clinical data registers [1, 2], attention is turning increasingly to special techniques for comparing hospitals [3–5].

The fundamental question when comparing the treatment effects observed in two independent patient cohorts is whether the two cohorts are comparable in relation to their structures and the intensity of the disorder. In non-randomised comparisons the observed relation between the endpoint and the care-provider may even be attributed to a third criterion (a so-called confounder). The groups treated in the various hospitals may vary widely in terms of the covariates observed because there is no experimental control over the hospitals to which patients are referred. If these

KARGER

Fax +49 761 452 07 14
E-mail Information@Karger.de
www.karger.com

Accessible online at:
www.karger.com/fkm

Dr. rer. biol. hum. W. Weidenhammer
Zentrum für naturheilkundliche Forschung
II. Med. Klinik und Poliklinik, Technische Universität München
Kaiserstr. 9, D-80801 München
Tel. +49 89 72 66 97-0, Fax -21
E-mail Wolfgang.Weidenhammer@lrz.tu-muenchen.de

covariates also have prognostic power with regards to the outcome of treatment, the difference between the treatment groups is distorted. In such a case the size of the observed difference is not equal to the true difference between treatment outcomes.

This situation necessitates the use of statistical correction or adjustment techniques that allow most reliable estimates of the treatment differences assuming that the patient groups are comparable.

The objective of this paper is to test two such methods for comparing the treatment outcomes in two hospitals. Both hospitals (referred to here as 'A' and 'B') co-operated with the Zentrum für naturheilkundliche Forschung (Centre for Complementary Medicine Research) within the framework of the 'Munich Model' hospital network at the time. The analysis was based on two prospective observational studies which were conducted in a virtually identical form and thus with equal documentation techniques in both hospitals. Both studies aimed to record the courses of migraine patients during and after inpatient therapy. As part of the quality assurance measures they described the quality of the short- and medium-term outcomes. Over a period of time set by the hospitals all patients referred for admission because of migraine (ICD-10: G43.0 or G43.1) were included in the observational study. There was no systematic selection.

The two hospitals' standard practice of giving appointments with the correspondingly long waiting times enabled the acquisition of pre-admission data. Documentation in the hospitals was restricted to the times 'Admission' and 'Discharge'. After their discharge the patients were sent by post questionnaires 3 time points (2, 6 and 12 months after discharge), which were identical each time, with the request to complete and return them to the hospitals. A 4-week headache diary was also used. All patients gave their informed consent to their anonymised data being forwarded for scientific analysis.

The present discussion of the results is limited to some indicators that are relevant for assessing headache therapy. The headache diaries were not evaluated because only a part of the patients had kept them. The before–after effects were exemplified by the study time points 'Discharge' and '6 Months after discharge'.

The diagnoses of migraine were verified by the criteria laid down by the IHS (International Headache Society). In one of the hospitals the diagnosis was confirmed previously by a specialist. The hospitalisation periods in both hospitals were comparable to an inpatient time of mostly 4 weeks. Exceptions were only made in individually substantiated cases. The treatment concepts will not be described in order to preserve the anonymity of the two hospitals. In any case, the concepts are of secondary importance for this analysis, which was conducted to examine the methodology.

Objectives

Are the patient groups in the two hospitals comparable? If not, which method of case-mix adjustment is recommended? What differences are revealed in the quality of the treatment outcomes in both hospitals?

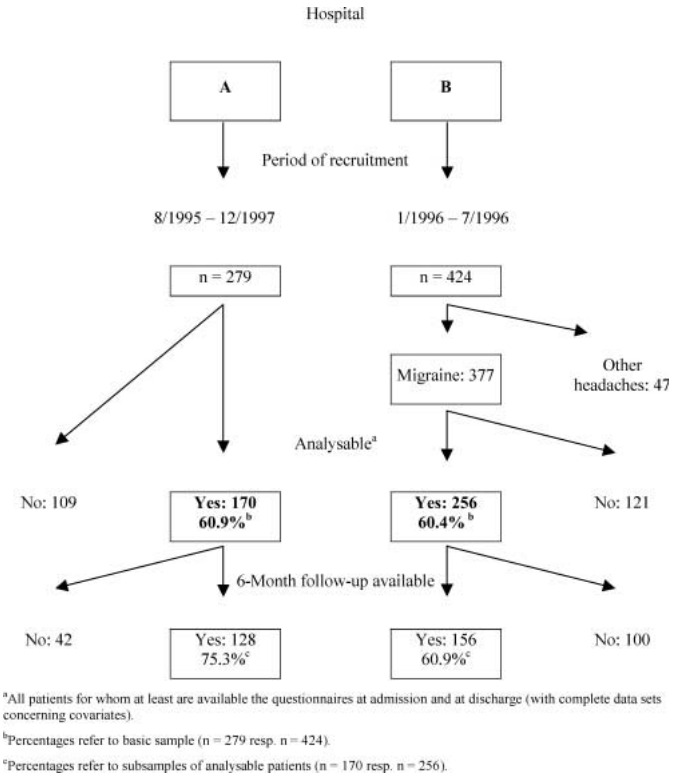

Fig. 1. Flow chart of numbers of patients being recruited and analysed.

Methods

Patients

Of the 703 headache patients originally documented in the two observational studies, 426 suffering from migraine or migraine and tension-type headache (according to the criteria of the International Headache Society) were selected for the secondary analysis. 170 were treated in hospital A, 256 in hospital B. The relative proportion of evaluated patients from the total patient sample was about 60% in both hospitals (fig. 1). The proportion of female patients in the two hospitals was 87%, with only minor deviations. The mean age of the patients was 45.5 ± 11.7 (range 14–73), which only varied marginally between the two patient groups.

The feedback rate of evaluable 6-months questionnaires was 75.3% for hospital A, and 60.9% for hospital B (difference statistically significant, $p < 0.01$).

Study Variables

The study variables to be analysed were decided on beforehand (table 1). They were selected according to the relevant recommendations for the conduct of clinical studies for describing the therapy outcomes for headache patients [6]. SF-36 [7] is a questionnaire for recording health-related quality of life. Treatment results were shown as the difference between the values recorded before and afterwards. The baseline values of these parameters were considered as potential covariates in all cases. They were supplemented by a number of basically descriptive sample parameters (e. g. age, duration of symptoms, subjective expectations of success). The variable 'data compliance' was included to cover the possible selective effect of incomplete feedback data (readiness to return questionnaires).

Statistics

General variance analysis models were used for the statistical analyses. Testing for differences between the patient samples from both clinics was performed with simple variance analysis with the variable 'hospital' as the be-

Table 1. Variables used for analysis in different functions

Variable	Definition	Used as		
		potential covariate	dependent variable at discharge	dependent variable 6 months after
Age	Years.	X		
Diagnosis	Migraine or migraine with additional headache of tension-type.	X		
Duration of complaint	Information by patient (Question: Since how many years do you suffer from headache?).	X		
Expected success	Patients rating before treatment (Question: Do you believe that you will profit from our treatment? Answers: 1 = yes definitely, 2 = yes may be, 3 = I don't know, 4 = I have doubts, 5 = not at all).	X		
Compliance in documentation	Completeness of data sets all over the course (number of examinations available: min. = 2 up to max. 6).	X		
Frequency of headache attacks	Frequency of migraine attacks in the previous month (retrospective estimation by patient).	X	X	X
Intensity of pain	Mean intensity of headache in the previous month, min. = 0 (no pain) up to max. 50 = extremely severe).	X	X	X
Concomitant symptoms	Sum score of patients ratings concerning five concomitant symptoms (Nausea, vomiting, photo-/phonophobia, other symptoms) graded as 0 = no, 1 = little, 2 = moderate, 3 = strong; Score ranges from 0 to max. 15.	X	X	X
Intensity of complaint	Intensity of main complaint, 100 mm-visual analogue scale (0 = no up to 100 = extremely strong).	X	X	X
Physical health	Sum score PCS (Physical Component Score) of SF-36; normal sample: mean = 50, SD = 10.	X		X
Mental health	Sum score MCS (Mental Component Score) of SF-36; normal sample: mean = 50, SD = 10	X		X
Goal attainment	Deviation of the initially expected intensity of complaint from that intensity achieved after treatment (difference of two VAS measures, negative values mean goal failed).		X	

tween-group factor. For analysing the differences between the hospitals regarding treatment outcomes, 3 models were calculated for each dependent variable:

1. Univariate variance analysis with the variable 'hospital' as the between-group factor.
2. Direct covariance analysis model: A number of potential covariates were included in the comparison of the parameters of the two patient groups. Classification factors and influencing variables (covariates) suppose to have a linear effect on the dependent variable (analysis of covariance I).
3. Compilation of the covariates into a so-called propensity score. Here, the propensity score estimates the probability of a specific patient belonging to a particular hospital depending on the given covariates [8]. In the case of two groups this probability is modelled by a logistic regression model [9]. In this instance the stepwise forward method was employed. Ultimately, the propensity score is used as the sole covariate in the covariance analysis (analysis of covariance II).

Results

Four of the total of 11 variables originally regarded as potential covariates showed statistically significant differences between the patient groups in the two hospitals. The patients in hospital B reported having their symptoms for longer than those in hospital A ($p < 0.01$). The mean number of headache attacks per month was higher in hospital B ($p < 0.05$). In hospital A more documented study times were available for the patients ($p < 0.05$), and their expectation that they would benefit from the hospital stay was considerably higher than among the patients in hospital B ($p < 0.001$). The proportion of patients who had tension headache in addition to migraine was slightly higher in hospital B ($p < 0.10$).

All the potential covariates were also entered in the logistic regression for calculating the propensity score. With an inclusion criterion of $p < 0.10$, 6 variables finally remained in the model (table 2). Their prognostic importance is given in the following descending order of importance: expectation of success, duration of symptoms, data compliance, attack frequency, mental health, and age.

The 20% percentiles (= quintiles) were determined for the propensity scores of all patients in both groups. The distribution of the propensity scores classified in this manner shows marked differences in both patient groups (fig. 2). If the propensity scores had been distributed evenly, then no covariance adjustment would have been necessary.

Solely on the basis of knowing the propensity score (= probability of belonging to hospital B), the correct forecasts were made for 42.9% of the patients in hospital A, and for 82.4% of the hospital B patients. Overall, 66.7% of the patients could be assigned correctly.

The results of the tests for differences between both patient groups concerning the treatment outcomes are shown in Table 3 (treat-

Table 2. Results of the logistic regression for estimating the propensity score (inclusion criterion $p < 0.10$)

Variable in the equation	Regression coefficient B	Standard error	Wald-Statistic	df	Sig.
Expected success	0.634	0.162	15.264	1	0.000
Duration of complaint	0.030	0.010	8.525	1	0.004
Compliance in documentation	–0.244	0.093	6.943	1	0.008
frequency of headache attacks	0.037	0.016	5.541	1	0.019
Mental health	0.021	0.010	4.368	1	0.037
Age	–0.018	0.011	2.912	1	0.088
Constant	*–0.415*	*0.773*	*0.289*	*1*	*0.591*

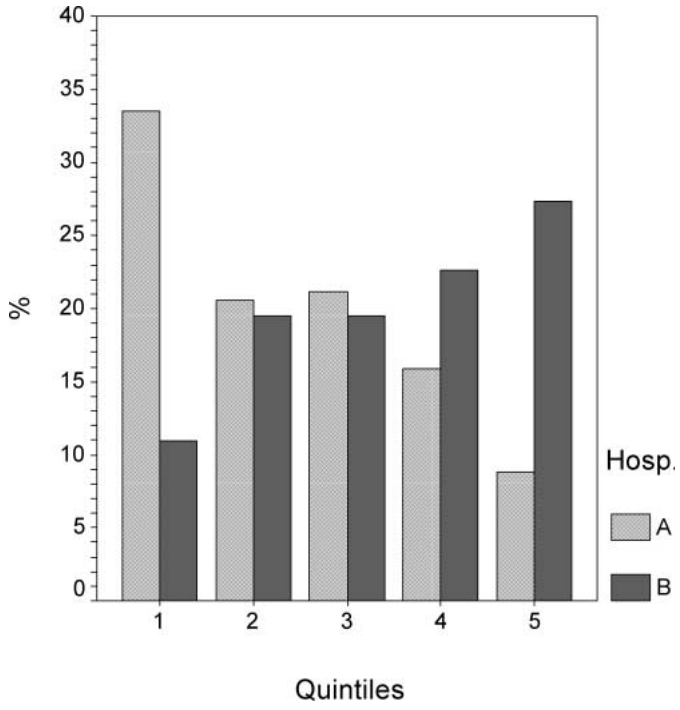

Fig. 2. Distribution of the propensity score (means probability for coming from hospital B) by quintiles in both patient groups.

ment effect at discharge). The upper part of the table shows the statistically descriptive results for the parameters as well as the results of the univariate tests for differences without regression analysis adjustment. The lower part of the table presents the results after adjustment by the two different analyses of covariance.

At the time of discharge, statistically significant differences were observed between the patient groups of both clinics with regard to both a reduction in the intensity of the disorders as well as the degree to which the expectations of reducing symptoms were met. Two other parameters (reduction in headache intensity and concomitant symptoms) showed trends towards different outcomes ($p < 0.10$).

All the covariates in analysis of covariance I are listed with a p value, i. e. only the parameters with a low value ($p < 0.05$) have a relevant influence on the dependent variable. This includes the p values for the tests for differences between the two patient groups after 'correcting' for the differences in the covariates. According to these, there were significant differences between the groups only with regard to the reduction in the intensity of the symptoms and headache as well as the degree to which expectations were met. The results of the variance analysis are essentially confirmed. The strongest effect of the covariance analysis is seen in the reduction of the frequency of headache attacks. After adjustment, the mean values of change in the two groups become closely approximated.

The results of analysis of covariance II, including the propensity scores, essentially agreed with those of analysis I. Here, too, the propensity score has the strongest influence with regard to changes in the frequency of the attacks.

6 months after discharge there were initially statistically significant differences between the patient groups in both hospitals with regard to a reduction in the frequency of attacks and improved quality of life. There was a trend towards different results regarding a reduction in the intensity of symptoms ($p < 0.10$).

After considering the direct influence of the covariates (analysis of covariance I), only 1 comparison test remains within the statistically significant range (increase in the SF-36 score for physical health). None of the other differences between the two groups are significant. Adjustment by the propensity score (analysis of covariance II) provides a similar result.

The different adjusting effects of the covariance analyses on the treatment outcomes are illustrated as an example. While there is no noticeable difference between the original and adjusted results for the reduction in symptoms at discharge (fig. 3, left), the reduction in the frequency of headache attacks after 6 months clearly shows the corrective effect (fig. 3, right). In both cases the difference between the two techniques of case-mix adjustment is minimal.

Discussion

In all cases where no generally valid standards are available for assessing the quality of results achieved by a care provider, attention concentrates on comparisons with the results of other care providers. Irrespective of whether a patient is looking for the 'best' hospital, a referring physician for the most promising treatment option, or hospital financiers looking for the most efficient investment for their financial resources, the decision is always influenced by comparisons. The reasons why a randomised study cannot always provide the most valid findings for this objective have been

Table 3. Comparison of both hospitals regarding the dependent variables (differences between admission and discharge from hospital) by simple ANOVA as well as with respect to the covariates (ANCOVA I and II)

	Freq. of headache attacks	Intensity of pain	Concomitant symptoms	Intensity of complaint	Goal attainment
Raw results					
Hosp. A					
Mean	2.74	3.86	1.62	31.85	–14.55
SD	4.83	10.28	3.69	32.46	26.81
N	163	155	154	162	156
Hosp. B					
Mean	3.79	5.75	2.28	41.08	–1.43
SD	8.07	9.18	3.60	25.06	20.42
N	240	250	251	251	242
Difference testing between hospitals (p-value)	0.137	0.055	0.074	*0.001*	*< 0.001*
ANCOVA I, *direct control for covariates (p-values)*					
Age	0.429	0.579	0.672	0.909	0.485
Diagnosis	0.851	0.471	0.516	0.192	0.080
Duration of complaint	0.477	0.159	0.420	0.614	0.740
Expected success	*0.012*	0.966	0.753	*0.014*	0.622
Compliance in documentation	0.116	0.431	0.728	*0.019*	0.160
Frequency of attacks	*< 0.001*	0.967	0.599	0.161	0.221
Intensity of pain	0.727	*< 0.001*	0.504	*0.033*	0.268
Concomitant symptoms	0.660	0.072	*< 0.001*	0.923	0.752
Intensity of complaint	0.327	0.313	0.692	*< 0.001*	0.454
Mental health	0.502	0.331	0.069	*0.025*	*0.016*
Physical health	*0.013*	*0.040*	*0.045*	*0.001*	0.059
Corrected difference testing between hospitals (p-value)	0.980	*0.019*	0.126	*< 0.001*	*< 0.001*
Adjusted for covariates					
Hosp. A, mean	3.36	3.61	1.69	31.29	–14.82
Hosp. B, mean	3.37	5.91	2.24	41.44	–1.26
ANCOVA II, *control for propensity score (p-values)*					
Propensity-score	*0.001*	0.357	0.690	0.219	0.409
Corrected difference testing between hospitals (p-value)	0.756	*0.035*	0.069	*0.001*	*< 0.001*
Adjusted for propensity score					
Hosp. A, mean	3.23	3.67	1.58	31.14	–14.96
Hosp. B, mean	3.45	5.87	2.30	41.54	–1.17

widely discussed [10, 11]. Amongst other things in the field of alternative medicine, one is often confronted with the situation in which randomization is not feasible and therefore decisions have basically to be made on the basis of non-randomised comparative studies. A number of methodological approaches are available for making such comparisons as fair as possible. The main objective is to increase the homogeneity in structural features of the patients, whenever necessary. This can be achieved by forming subgroups by pair or frequency matching [12]. Other techniques adjust the parameters statistically, usually based on linear regression models. The comparison between two groups of inpatients in hospitals providing treatment for chronic headache described here is an example of the latter approach.

The results show that the existing structural differences between the patient groups cast doubt on a direct comparison of the treatment outcomes. The application of variance analysis models using covariates, either by direct inclusion of multiple moderator variables or by taking the diversion to calculate a propensity score, improves the power of the results in a comparative study. In our case the two statistical techniques led to similar results, making the question of which method is more suitable a secondary one [13].

Instead, we consider the question concerning the nature of the relevant adjustment factors to be more important. In our case, in any evaluation of the effects change factors have, the baseline values usually dominate. Furthermore, it is not always clear which variables are relevant influencing factors. This clearly depends on a large number of context factors in a specific case. An at least partly standardised procedure would be welcomed to improve the transparency of the adjustment algorithm. In the case of assessing the effects of headache treatment this would contain a list of relevant covariates. At the same time, this restriction to a list of known confounders would prevent an 'exhaustive' procedure by which any number of alternative hypothetical, and unrecorded, influencing factors could be called upon if the result does not turn out as desired [14]. In the fields of alternative and complementary medicine it is difficult to reach a consensus on this point because of the large number of disease-unspecific patient characteristics (e.g. self-effectiveness, sense of coherence, constitution parameters) [15]. The

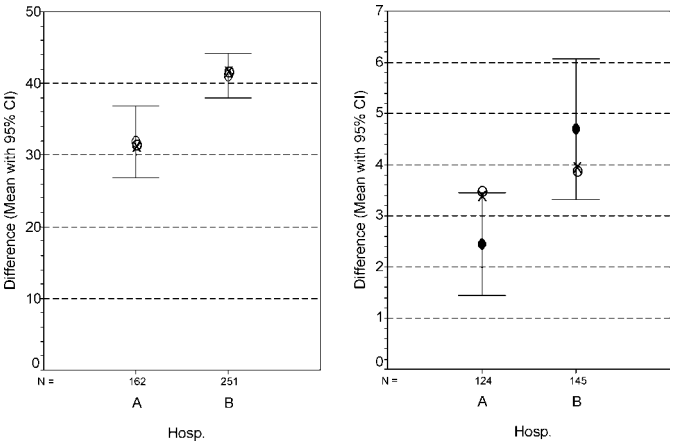

Fig. 3. Left: Decrease in intensity of complaint at discharge from hospital, comparison of both patient groups. Right: Decrease in frequency of headache attacks 6 months after discharge from hospital, comparison of both patient groups.
Bold circles = raw results without correction; O = adjusted according to ANCOVA I; X = adjusted according to ANCOVA II.

parameters are always selected in the shadow of a doubt as to whether the relevant variables have been taken into account, and whether an adequate form of measurability has been found.

More agreement is to be expected on the question of the relevant indicators of treatment outcome. Clear recommendations have been made here, at least in the conventional sphere [6]. In contrast to clinical studies conducted to test hypotheses, one should not restrict oneself to one single endpoint if possible when comparing the treatment outcomes achieved by different care providers. In such cases, presenting the comparison with various quality indicators in the form of a profile would simplify decision-making [4, 5].

This brings us full circle back to the original questions, and we must ask ourselves whether the results of the example presented here can answer them. Assuming that the statistically adjusted comparison between the two hospitals was 'fair' [16], can either of the two clinics be recommended from the viewpoint of the patients, the referring doctors, or the funding agencies? Any general conclusions must initially be treated with caution because the result of the comparison is not sufficiently clear. However, the comparison must contain sufficient information to simplify a decision reached by looking at the results from different angles and with different evaluation of the quality parameters, even if the decision is not unanimous.

Our example of the comparison between hospitals is intended as a contribution to this discussion, which becomes even more complex when comparing more than two hospitals. 'Ranking' a large number of care providers would be an increasingly demanded approach. Comparing each provider with a 'standard', i. e. the commonly required quality level of outcome, would also finally result in the problem of appropriate adjustment. For any valid decision this should be examined carefully, even though standard methods of case-mix adjustment do not guarantee removal of bias [17].

References

1 Fine LG, Keogh BE, Cretin S, Orlando M, Gould MM: How to evaluate and improve the quality and credibility of an outcomes database: validation and feedback study on the UK Cardiac Surgery Experience. BMJ 2003;326:25–28.
2 Motheral B, Brooks J, Clark MA, Crown WH, Davey P, Hutchins D, Martin BC, Stang P: A checklist for retrospective catabase studies – report of the ISPOR Task Force on retrospective databases. Value Health 2003;6:90–97.
3 Schubert-Fritschle G, Schmidt M, Klamert A, Hölzel D: Zur Variation der Basisdaten zwischen Kliniken in einem Tumorregister; in Michaelis J, Hommel G, Wellek S (ed): Europäische Perspektiven der Medizinischen Informatik, Biometrie und Epidemiologie. München, MMV Medizin Verlag, 1993.
4 Normand SL, Glickman ME, Gatsonis CA: Statistical methods for profiling providers of medical care: Issues and applications. J Am Stat Assoc 1997;92:803–814.
5 Schneeweis S, Sengha O: Weiterentwicklung von Krankenhausbetriebsvergleichen: Wie wichtig ist Risiko-Adjustierung für den Krankenhausvergleich? in Sieben G, Litsch M (eds): Krankenhausbetriebsvergleiche. Berlin, Springer, 2000.
6 International Headache Society Clinical Trials Subcommittee: Guidelines for controlled trials of drugs in migraine, ed. 2. Cephalalgia 2000;20: 765–786.
7 Bullinger M, Kirchberger I: SF-36 Fragebogen zum Gesundheitszustand. Göttingen, Hogrefe, 1998.
8 Rosenbaum PR, Rubin DB: The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55.
9 Breslow NE, Day NE: Statistical methods in cancer research – the analysis of case control studies. Lyon, IARC Scientific Publications, no. 32, 1980.
10 Abel U, Koch A (ed): Nonrandomized Comparative Clinical Studies. Düsseldorf, Symposion Publishing, 1998.
11 Radford MJ, Foody JM: How do observational studies expand the evidence base for therapy? JAMA 2001;286:1228–1230.
12 Trampisch HJ, Windeler J (ed): Medizinische Statistik. Berlin, Springer, 1997.
13 Lüdtke R, Weber U, Fischer I, Friese KH, Moeller H: An example on the value of non-randomisation in clinical tials in complementary medicine. Forsch Komplementärmed Klass Naturheilkd 2002;9:105–109.
14 Holzkamp K: Kritische Psychologie. Frankfurt, Fischer, 1972.
15 Melchart D, Weidenhammer W, Linde K, Saller R: 'Quality Profiling' for complementary medicine: The example of a hospital for traditional chinese medicine. J Altern Complement Med 2003;9:193–206.
16 Binder H, Frick U, Barta W: 'Fair ist, Gleiches mit Gleichem zu vergleichen?' Eine Simulationsstudie zu den Krankenhausvergleichen nach § 5 Bundespflegesatzverordnung. Poster, 1. Deutscher Kongress für Versorgungsforschung, 11.–12.10.2002, Köln.
17 Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, et al: Evaluating non-randomised intervention studies. Health Technol Assess 2003;7(27).