International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 17(S1): S65–S69 (2008)
Published online in Wiley InterScience
(www.interscience.wiley.com) **DOI**: 10.1002/mpr.251

Development and validation of a cluster-based classification system to facilitate treatment tailoring

SUSAN E. COLLINS, IRIS TORCHALLA, MARTINA SCHRÖTER, GERHARD BUCHKREMER, ANIL BATRA

University Hospital Tübingen, Tübingen, Germany.

Abstract

Aims: *The objectives of this study were to replicate smoker profiles identified in Batra et al. (in press) and to develop a cluster-based classification system to categorize new cases into smoker profiles so that an appropriate tailored intervention could be applied.*

Methods: *Participants were smokers in southwest Germany who sought treatment for smoking cessation. In the first sample, discriminant analysis was used to create classification formulas for a future study (classification sample: n = 165). The second sample served to replicate the smoker profiles, which included participants reporting symptoms of hyperactivity/novelty-seeking, depressivity or high nicotine dependence as well as participants scoring low across smoking and psychological variables (replication sample:* N = *134).*

Results: *Part 1 was focused on the development of formulas, using Fisher's coefficients, with which new cases could be classified. Part 2 adequately replicated previous findings concerning the smoker profiles, such that 70% of participants in the second sample were classified identically using cluster analysis and classification formulas.*

Conclusions: *The smoker profiles found in a previous study were replicated, and a classification system was developed for a future study which will test the efficacy of tailored treatments for the different smoker profiles. Copyright © 2008 John Wiley & Sons, Ltd.*

Key words: smoker profiles, smoker typology, cluster analysis, smoker classification, nicotine dependence

Introduction

Despite relatively high short-term abstinence rates of 50% to 80%, efficacy of pharmacobehavioural smoking cessation treatment typically decreases after treatment end, such that only 20–35% of participants remain abstinent one year following treatment (Cinciripini et al., 1996; Haas et al., 2004; Perkins et al., 2001). These relatively high rates of relapse indicate a need to develop new and more efficacious ways to help smokers maintain long-term abstinence.

One pathway to improving overall abstinence rates is to tailor interventions to address the needs of smokers who are at higher risk for relapse (see Ranney et al., 2006). In order to identify such at-risk smokers, a previous study was conducted to find theory-based smoker profiles that were associated with smoking relapse (Batra et al., in press). In this previous study, participants were classified into smoker profiles based on their responses to self-report questionnaires. The hypothesized profiles were identified using a *k*-means cluster analysis and included hyperactive/novelty-seeking, depressive, highly dependent, and low-scoring (i.e. smokers scoring comparatively low across smoking and psychological variables) profiles. Smoker profiles predicted both point-prevalence and continuous abstinence over the 12-month follow-up period (Batra et al., in press).

These profiles provided clinically meaningful information about smokers, and thus the first aim of the current study was to establish formulas based on the previous cluster analyses which could be used to classify new cases into smoker profiles. Given the small sample size in the previous study and the sample specificity of clustering techniques (Everitt et al., 2001), the second aim of this study was to replicate the previous cluster analysis in an additional sample. To this end, two classifications were compared: (a) classification using formulas based on the previous cluster analyses and (b) classification using new cluster analyses.

Methods

Participants in the current study consisted of two samples of smokers residing in southwest Germany who had completed a six-week smoking cessation programme (see Batra and Buchkremer, 2004; Batra et al., in press). The characteristics of the first sample have been described elsewhere (Batra et al., in press). In the second sample, 95% of participants identified as German nationals, 50.7% were women, and the average age was 46.3 years (SD = 11.08). At baseline, participants reported smoking an average of 23.6 cigarettes a day (SD = 9.43), evinced moderate levels of nicotine dependence (Fagerström Test of Nicotine Dependence [FTND]: mean, M = 5.27, SD = 2.17) and had made an average of 4.4 (SD = 6.69) previous quit attempts. According to the results of *t*-tests and chi-square tests, the two samples did not significantly differ on the above measures (all *p*s > 0.05), which indicated that the samples were comparable.

All participants ($N_{part1} = 165$; $N_{part2} = 134$) completed questionnaires assessing psychological and smoking characteristics. Two questionnaires measured depressivity, including the Beck Depression Inventory-German version (BDI; Beck and Steer, 1987; Hautzinger et al., 1994) and the Negative scale of the Inventory of Self-communication for Adults (ISE-N; Tönnies and Tausch, 1981). The Questionnaire on Smoking Urges (QSU; Müller et al., 2001; Tiffany and Drobes, 1991) and the Fagerström Test of Nicotine Dependence (FTND-G; Heatherton et al., 1991; Rumpf et al., 1995) were used to measure nicotine dependence and craving. Finally, three measures were used to assess hyperactivity and novelty-seeking: the Attention Deficit and Hyperactivity Checklist (ADHD; Heßlinger et al., 2004), the novelty-seeking scale of the Tridimensional Personality Questionnaire (TPQ-NS; Cloninger, 1987; Defeu et al., 1995) and the Behavioural Activation System scale of the BIS/BAS questionnaire (Carver and White, 1994; Strobel et al., 2001).

Prior to analyses, the resulting summary scores were inspected for outliers and normality. Exploratory data analyses indicated positively skewed distributions for BDI, ADHD and ISE-N, and these variables were subsequently transformed using square root (BDI, ADHD) and $\log_{10}$ (ISE-N) transformations. All other variables showed normal distributions as confirmed by plots and statistical tests of normality. Because the optimization technique used in this study was scale-sensitive and the indicators were scaled differently, continuous indicators were standardised before the analyses were conducted.

Results

Part 1

A discriminant analysis was conducted on Sample 1, in which the profiles designated by the cluster analysis conducted in Batra et al. (in press) served as the dependent variables, and the summary scores of the measures mentioned earlier served as indicators. The goals of this analysis were to better describe the profiles created by the previous cluster analysis in terms of their discriminant functions and to create classification formulas to assign new cases in Sample 2 to smoker profiles. Because the previous *k*-means cluster analysis aimed to minimise within-cluster variation and maximize between-cluster variation, inferential statistics reflecting model and predictor significance in the discriminant analysis were likely to be upwardly biased and are therefore not reported.

As shown in Table 1, loadings of predictors on discriminant functions indicated that BDI, ADHD and ISE-N scores discriminate between smokers with higher and lower depressive symptoms, FTND and QSU discriminate between smokers with higher and lower levels of nicotine dependence; and TPQ and BAS discriminate between smokers with higher and lower levels of novelty-seeking.

The discriminant analysis also yielded Fisher's coefficients which were applied to linear classification formulas and used to categorise new cases in Sample 2 into their respective smoker profiles (Tabachnick and Fidell, 2001). Classification scores (C_j) were given by

$$C_j = c_{j0} + c_{j\mathrm{FTND}}X_{\mathrm{FTND}} + c_{j\mathrm{QSU}}X_{\mathrm{QSU}} + c_{j\mathrm{BDI}}X_{\mathrm{BDI}} + c_{j\mathrm{ISE}}X_{\mathrm{ISE}} + c_{j\mathrm{ADHD}}X_{\mathrm{ADHD}} + c_{j\mathrm{BAS}}X_{\mathrm{BAS}} + c_{j\mathrm{TPQ}}X_{\mathrm{TPQ}} \quad (1)$$

DOI: 10.1002/mpr

Table 1. Loading matrix of indicators on functions

	Functions		
	Depressive	Nicotine dependent	Novelty-seeking
BDI	0.632[1]	−0.337	−0.015
ADHD	0.579[1]	−0.173	−0.023
ISE-N	0.515[1]	−0.357	−0.070
FTND	0.259	0.578[1]	−0.090
QSU	0.159	0.507[1]	−0.430
TPQ-NS	0.143	−0.062	0.803[1]
BAS	0.110	0.434	0.537[1]

Note: BDI, Beck Depression Inventory; ADHD, Attention Deficit and Hyperactivity Checklist; ISE-N, Negative scale of the Inventory of Self-communication for Adults; FTND, Fagerström Test of Nicotine Dependence; QSU, Questionnaire on Smoking Urges; TPQ, Tridimensional Personality Questionnaire; BAS, Behavioural Activation System scale of the BIS/BAS.
[1] Largest absolute correlation between each variable and any function.

where j = a given profile, c = the corresponding Fisher's coefficient from Sample 1, and X = the raw score on the specified predictor in Sample 2. Each case was assigned to the profile for which it had the highest classification score.

Part 2

The goal in Part 2 of this study was to replicate the smoker profiles and test the generalisability of the formulas created in Part 1. To achieve these ends, participants in Sample 2 were assigned to subgroups using two different methods: (a) the classification formulas derived in Part 1 and (b) a new *k*-means cluster analysis conducted on Sample 2 with four specified groups and iterations = 10.

The two partitioning results were compared to confirm cluster robustness (Everitt et al., 2001). Because the methods yielded an equal number of clusters, only a stringent exact reclassification was considered (Everitt et al., 2001). Results indicated that participants were consistently classified into their respective smoker profiles 70% of the time. Further, intermethod reliability was used to test the consistency of the two methods of classification, such that a higher consistency in classification yielded a higher Cohen's κ coefficient. According to cut-offs provided by Altman (1991), intermethod reliability analyses indicated adequate consistency of the two partitioning methods (κ = 0.60). A qualitative comparison of the profile scores on the seven indicators further confirmed cluster robustness (see Figure 1).

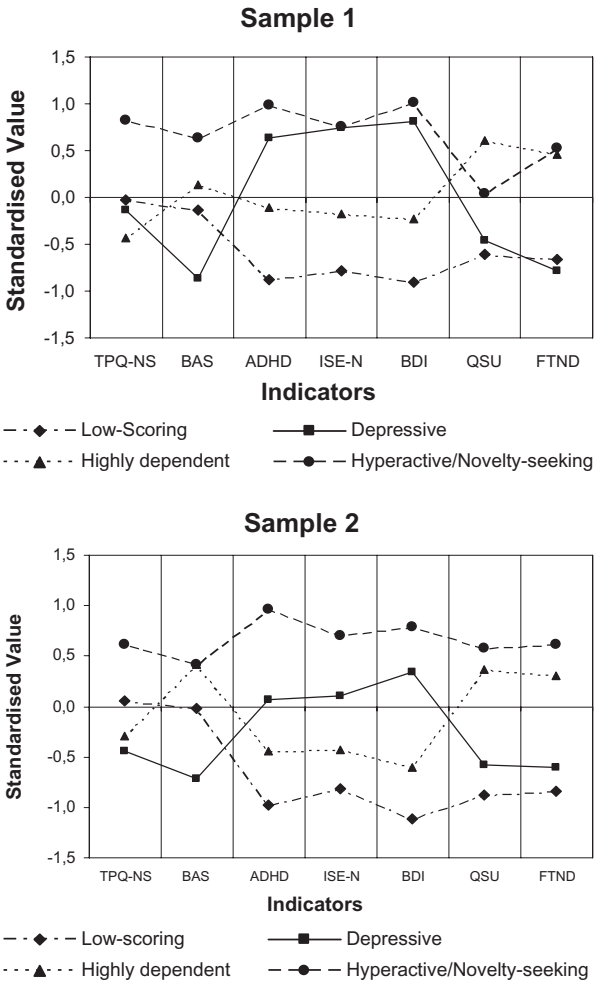

Figure 1. Comparison of indicator scores for smoker subgroups in study Samples 1 and 2. BDI, Beck Depression Inventory; ADHD, Attention Deficit and Hyperactivity Checklist; ISE-N, Negative scale of the Inventory of Self-communication for Adults; FTND, Fagerström Test of Nicotine Dependence; QSU, Questionnaire on Smoking Urges; TPQ, Tridimensional Personality Questionnaire; BAS, Behavioural Activation System scale of the BIS/BAS.

Discussion

In the current study, we developed and tested a system to classify new cases to smoker profiles. Findings indicated that this method, based on Fisher's classification

DOI: 10.1002/mpr

formulas from a previous study sample and self-report responses on psychological measures, is a feasible means of classifying new cases for the future treatment phase of the programme. Further, in spite of the relatively small sample size in the original study ($N = 165$), the current analysis provided evidence that the proposed smoker profiles are replicable and relatively robust.

Despite these promising findings, limitations of the current study warrant discussion. First, the multivariate analyses (i.e. *k*-means cluster analysis, discriminant analysis) used in the current study assume normality and are sensitive to outliers (Everitt et al., 2001; Tabachnick and Fidell, 2001). Fortunately, following transformation, the current data met the normality assumption and did not contain outliers. Further, cluster analyses may be subject to overinterpretation due to potential algorithm artefacts (Everitt et al., 2001). The current study, however, replicated profiles found in a previous study, which clearly reduced the risk of overinterpretation of the initial findings. Finally, a *k*-means cluster analysis was used in the current study to maintain consistency with the previous study analyses and thereby provide a replication of the methods as well as the results. It should be noted, however, that newer methods, such as latent class analysis (Muthen, 2002), may provide more flexible and elegant clustering possibilities. Although the current sample size would not have provided sufficient power to test the hypotheses using latent class analysis (Nylund et al., 2007), the follow-up study could incorporate this method as a further test of cluster robustness in a new, larger sample.

Despite its limitations, this study provided classification formulas to assign new cases to smoker profiles. Further, this study provided preliminary evidence for the robustness and validity of the smoker profiles in a new sample. An intervention study is being planned to serve as a further replication of the profiles and to establish the clinical utility of such profiles in tailoring pharmacobehavioural smoking cessation treatment to the needs of at-risk smokers.

Acknowledgements

This paper is an extended and modified version of an oral presentation at the BMBF Conference "Understanding Addiction: Mediators and Moderators of Behavior Change Processes", Dresden, May 25–26, 2006.

The authors would like to thank Sandra Eck, Sabine Konnerth, Matthias Pforr and Katrin Törpisch for the data management and entry on this project. This research was supported by grants from the Ministry of Science, Research and the Arts and the German Federal Ministry of Education and Research (01EB0411) to Professor Dr Anil Batra.

Declaration of interest statement

Anil Batra has accepted funding for smoking cessation research from Pfizer Consumer Health Care, Pharmacia & Upjohn, Pfizer GmbH, Sanofi Sythelabo, and GlaxoSmith-Kline in the past. However, the current study was not funded by any of these companies. Gerhard Buchkremer has not accepted funding for smoking cessation research in the past 20 years. All other authors declare that they have no conflicts of interest.

References

Altman DG. Practical Statistics for Medical Research. London: Chapman & Hall, 1991.

Batra A, Buchkremer G. Tabakabhängigkeit. Ein Leitfaden für Therapeuten [Tobacco Dependence: A Manual for Therapists]. Stuttgart: Kohlhammer, 2004.

Batra A, Collins SE, Torchalla I, Schöter M, Buchkremer G. Multidimensional smoker profiles and their prediction of smoking following a pharmacobehavioral intervention. J Subst Abuse Treat 2008; (In press) DOI: 10.1016/j.jsat.2007.08.006

Beck A, Steer R. Beck Depression Inventory – Manual. San Antonio, TX, USA: The Psychological Corporation, 1987.

Carver C, White T. Behavioral inhibition, behavioral activation and affective responses to impending reward and punishment: the BIS/BAS scales. J Person Soc Psychol 1994; 67: 319–33.

Cinciripini PM, Cinciripini LG, Wallfisch A, Haque W, Van Vunakis H. Behavior therapy and the transdermal nicotine patch: effects on cessation outcome, affect and coping. J Consult Clin Psychol 1996; 64: 314–23.

Cloninger C. A systematic method for clinical description and classification of personality variants: a proposal. Arch Gen Psychiatry 1987; 44: 573–88.

Defeu P, Kuhn S, Schmidt LG. Prüfung der Gütekriterien einer deutschen Version des Tridimensional Personality Questionnaire von Cloninger bei Alkoholabhängigen [Psychometric test of the German verison of Cloninger's Tridimensional Personality Questionnaire in alcohol dependent patients] Sucht 1995; 41: 395–407.

Everitt BS, Landau S, Leese M. Cluster analysis. New York: Oxford University Press, 2001.

Haas AL, Munoz RF, Humfleet GL, Reus VI, Hall SM. Influences of mood, depression history and treatment modality on outcomes in smoking cessation. J Consult Clin Psychol 2004; 72: 563–70.

Hautzinger M, Bailer M, Worall H, Keller F. Das Beck-Depressions-Inventar (BDI). Bern: Huber, 1994.

Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991; 86: 1119–27.

Heßlinger B, Philipsen A, Richter R. Psychotherapie der ADHS im Erwachsenalter: Ein Arbeitsbuch [Psychotherapy for Adult ADHD: A Workbook]. Göttingen: Hogrefe-Verlag, 2004.

Müller V, Mucha R, Ackermann K, Pauli P. Die Erfassung des Cravings bei Rauchen mit einer deutschen Version des 'Questionnaire on Smoking Urges' (QSU-G). Z Klin Psychol Psychother 2001; 30: 164–71.

Muthen BO. Beyond SEM: general latent variable modeling. Behaviormetrika 2002; 29: 81–117.

Nylund KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. Structural Equation Modeling 2007; 14: 535–69.

Perkins KA, Marcus MD, Levine MD, Delia DA, Miller A, Broge M, Ashcom J, Shiffman S. Cognitive-behavioral therapy to reduce weight concerns improves smoking cessation outcome in weight-concerned women. J Consult Clin Psychol 2001; 69: 604–13.

Ranney L, Melvin C, Lux L, McClain E, Lohr KN. Systematic review: smoking cessation intervention strategies for adults and adults in special populations. Ann Inter Med 2006; 145.

Rumpf HJ, Hapke U, John U. Deutsche Übersetzung des Fagerström Test for Nicotine Dependence Greifswald: Institut für Epidemiologie und Sozialmedizin, 1995.

Strobel A, Beauducel A, Debener S, Brocke B. Eine deutschsprachige Version des BIS/BAS-Fragebogens von Carver und White. Zeitschrift Differentielle Diagnostische Psychologie 2001; 22: 216–27.

Tabachnick BG, Fidell LS. Using Multivariate Statistics. Boston, MA: Allyn and Bacon, 2001.

Tiffany ST, Drobes DJ. The development and initial validation of a questionnaire on smoking urges. Br J Addict 1991; 86: 1467–76.

Tönnies S, Tausch R. Inventar zur Selbstkommunikation für Erwachsene (ISE) [Self-communication inventory for adults ISE]. Z Klin Psychol 1981; 10: 52–63.

Correspondence: Anil Batra, Universitätsklinikum Tübingen, Osianderstr. 24, 72076 Tübingen, Germany.
Telephone (+49) 07071/29-82685
Fax (+49) 07071/29-4141
Email: anil.batra@med.uni-tuebingen.de