Published in final edited form as:

Addiction. 2010 December ; 105(Suppl 1): 11–12. doi:10.1111/j.1360-0443.2010.03207.x.

Latent class and mixture models' potential contributions to understanding connections between menthol and other cigarette smoking characteristics

BRIAN P. FLAHERTY

University of Washington, Department of Psychology, Box 351525, Seattle, WA 98195, USA

BRIAN P. FLAHERTY: bxf4@u.washington.edu

Keywords

Latent class; mixture model; psychometrics; cigarette smoking patterns; menthol

Recent advances in computation have greatly broadened the range of statistical models researchers may consider for their work. To the extent that a statistical model is a possible representation of phenomena of interest and a possible representation of the data-generating mechanism, this expansion of the range of specifiable models correspondingly opens new opportunities for theory development, modeling and testing. That is, new statistical models may lead to new research questions. Scientific areas that have proven difficult to clarify with standard approaches may be well suited for examination with new methodologies. The role of menthol in tobacco use and associated outcomes is one such domain [1]. This commentary introduces a large class of statistical approaches that fit well with existing models of tobacco use, yet also introduce important flexibility, potentially leading to a clearer understanding of cigarette smoking and menthol's role therein.

Standard analytical models such as regression or analysis of variance rely upon multiple assumptions, one of which is that the model applies equally well to everyone in the sample. That is, when a single regression model is estimated, it is assumed that the model and estimates describe everyone in the data (and population, if one is generalizing). Subsequent discussion often involves the authors' extrapolating from sample-based results to individual level processes. For example, finding a positive association between menthol preference and nicotine dependence in a sample could lead to a discussion about menthol exacerbating dependence symptoms at the individual level. This might then be followed by an effort to remove menthol from tobacco products with the hope that it leads to reduced levels of nicotine dependence, greater cessation success and better health outcomes. The association between menthol and nicotine dependence is on the sample level, but the theoretically oriented discussion and potential intervention strategy are at the individual level.

Historically, this single population model assumption was untestable. However, a broad class of statistical models allowing one to examine this assumption has and continues to be developed. Latent class and mixture models (for brevity, latent class and mixture will be abbreviated as LC) [2–4] are statistical models that allow qualitative differences among population subgroups (i.e. the latent classes or mixture components). Essentially, the LCs

capture different patterns of association among measured items. Rather than assuming that menthol preference and nicotine dependence are associated positively among all members of a population, one could instead test a model with two subgroups: one in which menthol preference and nicotine dependence are associated positively and another in which they are unrelated.

If the latter situation were true, i.e. two different populations with different associations between menthol preference and nicotine dependence, and that class structure is ignored, then the sample-based parameter estimates are unlikely to represent anyone. Sample averaged estimates would, in fact, represent the overall sample, but they would only be useful for describing the sample. They would not help to clarify the connection between menthol and an individual's nicotine dependence. In fact, estimates based upon models that ignore class structure may be quite misleading. It depends upon the size of the classes and differences in associations across the classes.

If the two classes in this example were the same size in the population (50% each), then the sample average estimates are the average of the two within-class associations. Conversely, if 90% of the sample were from class 1 and 10% from class 2, the sample average estimates would be much closer to the class 1 association; however, it would probably be attenuated. In any case, ignoring the class structure may make accurate substantive interpretation far more difficult.

Tobacco researchers have long discussed different types of smokers. Class or categorical models of tobacco use progression are popular [5]. Different types of smokers have also been discussed in the adult smoking literature [6–9]. The notion of qualitative differences underlies these group labels. For example, light smokers are often discussed in terms qualitatively different from regular or stereotypical smokers: non-dependent, low-rate, driven by external factors, social smokers, etc. The patterns and associations among items imposed by qualitatively different types of smokers maps clearly onto the LC statistical model.

In addition to these different types of smokers, group differences in smoking have been identified across demographic and other social categories [10]. For example, African American and Hispanic smokers have historically had lower smoking rates and lower life-time consumption. Many of these same demographic and social groups also experience tobacco-related health disparities [11]. This combination of lower typical use and relatively poorer health outcomes has led to some of the interest and research on menthol. For example, African Americans are more likely to prefer mentholated cigarettes [12], and these smokers also experience some of the worst tobacco-related health disparities [13]. However, these differences among demographic and social groups raise questions about the structure of smoking within these groups. Do the different types of smokers mentioned above look the same across demographic groups? Are there simply fewer heavy smokers among African Americans (for example), or does the profile of heavy smokers among African Americans look different from the stereotypical profile of heavy smokers? Are light smokers similar across demographic groups too, or do they vary? Finally, does menthol play the same role in smoking across demographic groups and qualitatively different types of smokers?

LC models conform well to the research situation described. Qualitatively different types are discussed, yet a precise description of these groups is unavailable. In LC models, unlike discriminant function analysis or logistic regression, subgroup membership is unobservable, hence the 'latent' in latent class. LC models may be used for either exploratory or confirmatory analysis, similar to factor analysis [14]. In the absence of theory indicating how many and what classes to expect, LC analysis is exploratory. What are the prevalent

cigarette smoking patterns? Are they comparable across demographic and other social groups? If not, similarities and differences in these smoking types may be important theoretically. LC models may also be used as psychometric latent variable models, paralleling how the factor model is used as a psychometric model for continuous latent dimensions [15]. This is a necessary step in developing a replicable approach to measuring smoking patterns. As knowledge and theory accrue, theories may be tested in a confirmatory manner. Does a parallel set of classes hold across demographic groups? Do the patterns of effects implied by a particular theory hold for tobacco classes in other demographic groups?

LC models may provide a more refined characterization of tobacco use, empirically identifying common patterns between and within populations. If, in fact, associations between covariates and outcomes differ across demographic groups and/or smoking patterns, LC models are a natural step forward. Researchers interested in these models should consult the references included here and other peer-reviewed literature in order to familiarize themselves with model use and performance. Free programs available to estimate these models include PROC LCA and PROC LTA for SAS (<http://methodology.psu.edu/>), poLCA for R (<http://userwww.service.emory.edu/~dlinzer/poLCA/>) and GLLAMM for STATA (<http://www.gllamm.org/>). (These are listed in the author's opinion of their ease of use.) Commercial software options include Mplus and Latent-Gold. These models provide a rich and flexible framework for description, as well as theory-building research.

Acknowledgments

This work was supported by NIH grant R37DA18673.

References

1. Hebert JR. Invited commentary. Menthol cigarettes and risk of lung cancer. *Am J Epidemiol.* 2003; 158:617–20.
2. McCutcheon, AL. *Latent Class Analysis*. Thousand Oaks, CA: Sage; 1987.
3. Clogg, CC. Latent class models. In: Arminger, G.; Clogg, CC.; Sobel, ME., editors. *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. New York: Plenum Press; 1995. p. 311–59.
4. McLachlan, G.; Peel, D. *Finite Mixture Models*. New York, NY: Wiley; 2000.
5. Flay, B. Youth tobacco use: risks, patterns, and control. In: Orleans, CT.; Slade, JD., editors. *Nicotine Addiction: Principles and Management*. New York, NY: Oxford University Press; 1993. p. 365–84.
6. Shiffman S. Tobacco ‘chippers’—individual differences in tobacco dependence. *Psychopharmacology (Berl)*. 1989; 97:539–47. [PubMed: 2498951]
7. Wortley PM, Husten CG, Trosclair A, Chrismon J, Pederson LL. Nondaily smokers: a descriptive analysis. *Nicotine Tob Res.* 2003; 5:755–9. [PubMed: 14577992]
8. Hyland A, Rezaishiraz H, Bauer J, Giovino GA, Cummings KM. Characteristics of low-level smokers. *Nicotine Tob Res.* 2005; 7:461–8. [PubMed: 16085514]
9. Zhu SH, Pulvers K, Zhuang Y. Báezconde-Garbanati L. Most Latino smokers in California are low-frequency smokers. *Addiction.* 2007; 102:104–11. [PubMed: 17850620]
10. US Department of Health and Human Services. *Tobacco Use Among US Racial/Ethnic Minority Groups*. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease and Health Prevention and Health Promotion, Office on Smoking and Health; 1998.
11. Flaherty, BP.; Clayton, RR.; Alexander, LA., editors. *Addiction*. Vol. 102. 2007. Conceptual and methodological issues for research on tobacco-related health disparities.
12. Giovino GA, Sidney S, Gfroerer JC, O'Malley PM, Allen JA, Richter PA, et al. Epidemiology of menthol cigarette use. *Nicotine Tob Res.* 2004; 6:S67–S81. [PubMed: 14982710]

13. Fagan P, Moolchan ET, Lawrence D, Fernander A, Ponder PK. Identifying health disparities across the tobacco continuum. *Addiction*. 2007; 102(Suppl 2):5–29. [PubMed: 17850611]
14. Comrey, AL.; Lee, HB. *A First Course in Factor Analysis*. 2. Hillsdale, NJ: Lawrence Erlbaum Associates; 1992.
15. Flaherty BP. Assessing reliability of categorical substance use measures with latent class analysis. *Drug Alcohol Depend*. 2002; 68:S7–20. [PubMed: 12324172]