Addiction (1996) 91 (Supplement), S139–S145

COMMENTARY ON REPLICATIONS OF MARLATT'S TAXONOMY

Is Marlatt's relapse taxonomy reliable or valid?

RONALD M. KADDEN

University of Connecticut School of Medicine, Farmington, USA

Abstract

The Relapse Replication and Extension Project (RREP) has failed to provide empirical support for Marlatt's relapse taxonomy. Neither the reliability of the original Marlatt coding system nor its predictive or construct validity was supported by this group of studies. The present commentary explores a number of possible reasons for the generally negative outcomes. These findings should certainly lead to a re-evaluation of Marlatt's relapse taxonomy and its operationalization. Nevertheless, despite the negative results, there are a number of reasons why the general Relapse Prevention concept is likely to survive in some form: it has been widely adopted and imitated clinically, key elements of the taxonomy are often focal points of treatment, and clinical research studies have repeatedly supported some elements of the taxonomy (e.g. negative emotional states, social pressure, interpersonal conflict, positive emotional states and temptations/urges). The RREP also evaluated some modifications of the original taxonomy as well as the use of more structured assessment instruments, and some of these provided more promising results. Further developments will need to take into account both research needs for greater precision, most likely through the use of more structured assessment instruments, and clinical needs for richness of detail and sensitivity to a wide variety of life circumstances.

The concept of Relapse Prevention has become fairly popular in the clinical community. Hence the failure to provide empirical support for Marlatt's relapse taxonomy should be a matter of some concern to those who employ it as the basis for clinical assessment and intervention. Results of the Relapse Replication and Extension Project (RREP) provide a reason for pausing to reassess the concept and its constituent elements. However, despite the problems raised by the present set of studies, and the likely need for some revisions, there are nevertheless a number of reasons why the general approach may well survive a re-evaluation and continue to be influential.

In the first place, the Marlatt taxonomy has been widely imitated and applied, in a variety of forms, in clinical settings. It has struck a responsive chord that apparently is consistent with the experience of many clinicians. Although the imitations may not be exact replications of the original model, and may introduce some different elements, the general idea of examining prior relapse episodes as a starting point for clinical intervention has become popular. Marlatt's work has provided the impetus, and his overall con-

Correspondence to: Ronald M. Kadden, PhD, Department of Psychiatry, University of Connecticut School of Medicine, Farmington, CT 06030-2103, USA. Tel: +1 (860) 679 4249. Fax: +1 (860) 679 4077.

Submitted 29th March 1996; initial review completed 10th June 1996; final version accepted 2nd August 1996.

0965–2140/96/12S139–07 $8.00

Carfax Publishing Company

ceptualization of relapse has been central in these various efforts; it seems that many of the present investigators would concur in this observation (e.g. Longabaugh *et al.*, this issue).

Secondly, a number of the categories in "Level 2" of Marlatt's taxonomy make good clinical sense and are validated in treatment programs on a daily basis. Clients commonly identify negative emotional states, especially anger, frustration, depression, loneliness and anxiety, as having been antecedents to prior relapses and view them as high risk for the future. Clinicians also often view them as important and tend to make them a major focus of intervention. In addition, difficulties in coping with interpersonal conflicts and with various forms of social pressure are fairly commonly identified in the clinic as problems among recovering people. Some of the other elements in Marlatt's taxonomy are reported with less consistency and, as noted in the papers here, there is less agreement regarding their utility as relapse categories. Overall, however, the model provides a very convenient framework for conceptualizing the most frequently reported relapse episodes and for structuring clinical interventions. Furthermore, several of the categories of relapse precipitants are commonly reported by clients recovering from a variety of excessive appetites, including alcoholism, addiction to various drugs, smoking, gambling, and even among dieters.

Many rehabilitation programs have incorporated treatment elements (e.g., training in anger management, assertiveness, abstinence-orientated leisure activities, etc.) that address various of the relapse antecedents identified by the taxonomy. In the case of our own clinical program, the Marlatt taxonomy was of considerable help in identifying and labeling what our staff believe to be the central clinical issues, and providing a framework for conceptualizing and organizing our interventions. Certainly Marlatt has achieved his original goal of introducing basic psychological principles into addictions treatment in general, and specifically into the understanding of relapse (Marlatt, this issue).

Beyond clinical anecdotes, a third factor relates to the amount of evidence that has accumulated in the literature supporting at least some elements of the taxonomy. Of particular note is the frequency with which negative emotional states, defined and assessed in various ways, are identified as relapse precipitants (e.g. Chaney, O'Leary & Marlatt, 1978; Pickens *et al.*, 1985; Brownell *et al.*, 1986; Annis & Davis, 1988; Heather, Stallard & Tebbutt, 1991; Hodgins, el-Guebaly & Armstrong, 1995). Principal components analysis (Cannon *et al.*, 1990) and principal factor analysis (Isenhart, 1991) of the Inventory of Drinking Situations (IDS; Annis, 1982), a questionnaire based on the eight categories of the Marlatt model, both identified negative emotional states as a major factor. Heather & Stallard (1989) concur in the importance of negative emotional states, but view their role somewhat differently, as conditioned cues which elicit a conditioned response that is interpreted as craving. (For them, craving mediates the effects of most relapse precipitants.) Hodgins *et al.* (1995) conducted a study that included both prospective and retrospective descriptions of relapse antecedents, and found that clients' retrospective reports of relapse situations were not significantly influenced by a negative attributional bias that might have inflated the reporting of negative emotions (although Zywiak *et al.* (this issue) interpret the Hodgins *et al.* negative finding at the $p = 0.12$ significance level to be a "statistical trend", so that negative bias cannot be definitively ruled out). Social pressure has also been fairly commonly identified as a relapse precipitant, although somewhat less often than negative emotions (Chaney *et al.*, 1978; Heather *et al.*, 1991; Hodgins *et al.*, 1995).

Sandahl (1984) translated the Marlatt & Gordon (1980) questions into Swedish and generally confirmed their findings, with negative emotional states as the most common relapse precipitant, followed by interpersonal conflict, and social pressure. Litman *et al.* (1983) developed the Relapse Precipitants Inventory in Great Britain, and found that "unpleasant mood states" were a major factor, along with "external events and euphoria" (similar in part to Marlatt's "positive emotional states"), as well as an additional factor which Litman *et al.* characterized as "lessened cognitive vigilance".

The literature thus provides considerable evidence for the importance of negative emotional states which appear over and over again as relapse precipitants in a number of studies, as well as evidence for some other precipitants: social pressure, conflict, positive emotions, and temptations/urges. Certainly these demonstrations of the replicability of broadly defined relapse categories do not constitute verification of the

reliability or validity of Marlatt's taxonomy. Nevertheless, they do tend to support some of the categories that have been utilized, indicating their clinical utility and the likelihood of their survival, in one form or another, regardless of the fate of some of the other categories or of particular assessment instruments.

Many of the studies that provided these supportive findings involved the development of a variety of alternative scales and assessment methods to analyze relapses. As a result, there has been very little direct testing of Marlatt's original coding system—hence the present series of studies.

Longabaugh *et al.*'s (this issue) examination of Marlatt's coding system does not support its reliability. The best across-site agreement was 81%, for the negative emotional states category. Given the common scoring protocol that was employed across sites, the training given the raters, and the use of consensus ratings from each site, Longabaugh *et al.* concluded that the upper limit of inter-site agreement using Marlatt's coding system has probably been achieved.

It is unclear why the interrater reliabilities in the Longabaugh *et al.* study were low compared with reported values of 88% (Marlatt & Gordon, 1980) and 91.7% (Hodgins *et al.*, 1995) obtained in two other studies of the Marlatt coding system. However, Heather *et al.* (1991) found a lower rate of agreement (63%). In the Longabaugh *et al.* study, not all the raters were provided with the same information (one site used transcripts of clients' actual responses and the other two used notes made by the interviewers), although it is doubtful that this minor procedural difference alone could account for the reliability findings. An experience factor may also have been operative, since raters at the various sites ranged from undergraduate students to PhDs, and one site had considerable rater turnover, utilizing five different raters through the course of the study. It is unclear what impact varying rater experience may have had, but it should have been mitigated to some degree by the training that the raters received. However, there were reportedly some between-site variations in training procedures for the raters, and the actual amount of training given was not specified. Raters were only required to reach a level of 70% agreement with their trainers, and Longabaugh *et al.* acknowledge that this is below the usually acceptable minimum of 80%. The original training criterion of 80% had to be adjusted downward because it could not be achieved consistently, although it is not clear why this was such a problem, especially in view of the previously noted reports of reliability levels above 85% agreement (Marlatt & Gordon, 1980; Hodgins *et al.*, 1995).

Through the course of the Longabaugh *et al.* study, the overall agreement level was maintained above 70% at periodic checks of inter-rater agreement, but percentage of agreement for some of the specific Level 2 relapse categories, at some sites, was well below the 70% level. This may be an indication that computing reliability across three different research sites had a deleterious effect on reliability estimates, in comparison with studies that utilized only a single site. Furthermore, reliability estimates were related to category frequency: the more frequent a category, the greater its reliability (i.e. the negative affect category). Overall, the values for Level 2 agreement across sites were consistently superior to the Level 3 agreement, perhaps suggesting Level 2 as the appropriate starting point for further research on the Marlatt taxonomy.

With respect to validity, the findings of Maisto *et al.* (this issue) do not support the ability of the Marlatt taxonomy to predict the first relapse following treatment. Establishing predictive validity is an important step towards verifying clinical utility, even though the taxonomy in its original form may have been designed to describe, rather than predict, relapse episodes. Nevertheless, the taxonomy should have predictive validity if relapse *prevention* is to be a clinically useful concept; if none of the relapse antecedents that it identified were recurrent events, then treatment efforts that focus on those antecedents would be of little use. However, not every relapse antecedent is necessarily a clinically significant relapse precipitant, so the most recent antecedent, prior to the last relapse before treatment, may not have been the best one to focus on since it may not have been a clinically important one. A more critical or more typical incident may have actually occurred somewhat earlier in time, precipitating a chain of events, of which the relapse antecedent identified in the present study may have only been one, perhaps minor, step. In an alternative conceptualization, there may be no single crucial antecedent event at all, but rather a number of events in combination that affect the likelihood of relapse for a particular individual.

In the present research, one arbitrarily selected

variable that may have contributed to the lack of predictive validity was the length of prior abstinence required before a relapse could be said to have occurred. A relapse was defined as drinking that followed at least 4 days of abstinence, a parameter that could have been assigned a variety of different values, with potential impact on the findings. Someone with only 4 days of abstinence may have very different relapse vulnerabilities than other participants in the study who had more sustained periods of sobriety. In addition, the 4 days of abstinence and subsequent relapse may have occurred long before the interview at which the data were collected, in some cases a year or more in the past. With such variability in the timing of the baseline relapse incident, there may have been substantial differences in the accuracy of recall and in the number of intervening events that may have occurred.

It is also likely that the treatment which intervened between the baseline (pretreatment) relapse and the first post-treatment drinking episode differentially affected some relapse categories, presumably making them less likely to occur. Thus, the desired impact of treatment, in reducing the frequency of certain relapse categories, perhaps considerably so, may have had a negative effect on predictive validity from pre- to post-treatment. Stout *et al.* (this issue) conducted an analysis comparing the baseline pretreatment relapse and the first post-treatment relapse as well as the first two relapses that occurred post-treatment, providing some, albeit weak, support for this speculation.

Maisto *et al.* attempted to improve upon the predictive validity findings they obtained with the 8-category Marlatt coding system by collapsing those codings into the three categories that had been derived by Cannon *et al.* (1990) in their principal components analysis of the IDS. That three-component solution, composed of negative feelings, positive feelings and testing personal control scales, was similar to the three categories that Litman *et al.* (1983) had derived by principal components analysis of their Relapse Precipitants Inventory. Despite the consistency of the three derived categories across two independent studies, their use in the present instance did not result in improved predictive validity over that based on all eight categories of the Marlatt coding system. It is likely that some of the preceding comments with respect to the original predictive validity analysis apply as well to this re-analysis of the data.

Maisto *et al.* also used the Cannon *et al.* (1990) scoring system to compare Marlatt codings to IDS codings. In this instance, both sets of codings were redefined according to the Cannon *et al.* three-category system. Following this transformation, a significant relationship was found between the rescored Marlatt and IDS assessments of the baseline relapse episodes. This finding of concurrent validity is one of the few analyses in the present series of studies that is supportive of the Marlatt taxonomy.

However, the findings for construct validity were not supportive. The relationships of the Marlatt negative emotions and positive emotions categories to DSM-III-R diagnoses of various anxiety and affective disorders were not significant. It should be noted, however, that inasmuch as DSM-III-R was designed for formulating psychiatric diagnoses it may not have been well suited to the current purpose of identifying relapse precipitants. There may rather be anxiety or affective symptoms in addition to those required for a diagnosis, or perhaps only a subset of those required for a diagnosis, that serve as effective precursors to drinking. Perhaps comparison variables that utilized continuous scales of measurement, rather than categorical diagnoses, would have provided more appropriate comparison criteria as well as greater statistical power.

In addition to Maisto *et al.*, another paper in this issue (Stout *et al.*) also examined predictive validity, both of the Marlatt taxonomy and of another relapse coding system developed for this purpose. Their new coding system was based on Marlatt's, but added some new relapse categories, ratings of intensity, time from antecedent event to relapse, and information about the setting in which the relapse occurred. This study found that the original Marlatt taxonomy did not predict whether a post-treatment relapse would occur and, if one did occur, the pretreatment relapse category predicted neither the post-treatment relapse category nor the frequency or intensity of post-treatment drinking. The alternative relapse coding system did improve somewhat on the poor predictive validity of the Marlatt method, but not substantially. Given that there was at least some improvement, the authors concluded that their revised coding method was superior to Marlatt's

hierarchical taxonomy, although evaluation of its practical utility must await further assessment.

Zywiak *et al.* (this issue) conducted a factor analysis on data obtained with the Heather *et al.* (1991) questionnaire, which rates the relative importance of each of the Level 3 Marlatt relapse categories. The analysis identified negative emotions as the first factor, and social pressure along with positive emotions as the second factor, consistent with Cannon *et al.*'s (1990) principal components analysis of the IDS. (The third factor in the Zywiak *et al.* analysis was more complex than the third factor, testing personal control, identified by Cannon *et al.*). Zywiak *et al.* concluded that the importance of Marlatt's relapse categories, including the Level 3 categories, was supported by the fact that each one of them loaded on just one of the three derived factors. In addition, Zywiak *et al.* found a relationship between their first factor, negative emotions and independent measures of depression and anger, suggesting construct validity. Finally, there were significant correlations between the category ratings for the first and second relapses following treatment, providing an indication of predictive validity for these derived factors that eluded Maisto *et al.* with Marlatt's categories and with the three-component coding system that had been derived from the IDS. Perhaps the critical element in the Zywiak *et al.* finding was that both the predictor and subsequent relapses occurred *following* treatment, so that treatment was not a major intervening factor *between* the two relapse episodes as in Maisto *et al.*

With only a few exceptions, the findings reported in the present series of papers have not been supportive of the original Marlatt taxonomy, although some of the derivative systems have provided more promising results. Longabaugh *et al.* (this issue) made a number of suggestions that might improve upon the Marlatt coding system. They recommended increasing the structure of the interview used to identify relapse precipitants, reorganizing the relapse categories, modifying some of the more arbitrary scoring rules and attempting to identify relative degrees of validity of the various relapse-situation categories. These appear to be quite reasonable suggestions that ought to be pursued. Longabaugh *et al.* also suggested expanding the model to include more pervasive factors, so that the events immediately preceding a relapse would be only one element of a much larger picture. The broader picture might include client trait factors such as motivation, chronic negative affect and personality disorders, as well as environmental factors such as family relationships, neighborhood environment and social support network. These factors conceivably could be significant relapse precipitants in and of themselves, or might possibly moderate the impact of the more immediate relapse antecedents that have been considered heretofore.

Some researchers have recommended the option of allowing classification of relapse antecedents into more than just one of Marlatt's categories (e.g. Hodgins *et al.*, 1995), and Heather & Stallard (1989) provide a heuristic model that accommodates multiple relapse determinants. However, in a test of the proposal to include multiple relapse antecedents, Stout *et al.* (this issue) did not find substantial improvement over the single-category method. Taking this idea a step further, it may be fruitful to also look further back in time, prior to the most recent pre-relapse events (Hodgins *et al.*, 1995). In this regard, Litman *et al.* (1983) focused on the *number* of relapse precipitants that may be operating in a situation, a factor that is explicitly excluded in the Marlatt formulation. Possibly also, patterns of relapse episodes, frequency of occurrence of certain categories, and clients' ratings of the relative importance of various antecedent events in the relapse process (Heather *et al.*, 1991) should be taken into account. Annis & Davis' (1989) IDS "profiles" characterize a client's typical relapse situations or events over an extended time period, and as such may prove to be useful clinically (see also Annis & Graham, 1995).

Finally, clustering techniques may be helpful in analyzing relapse data and sorting out the complex possibilities enumerated above, as demonstrated by Baer & Lichtenstein (1988). They applied cluster-analytic techniques in studying relapses to smoking, and found clusters that differentiated situations that were negative, non-social and stressful from situations that were social and positive. They also found continuity across time, inasmuch as cluster membership was related to situational characteristics of prior relapse episodes.

Efforts to understand the relapse process may have substantial payoff, not only in terms of clinical care focused on particular high risk relapse categories, but also if it turns out that

relapses make a positive contribution to the recovery process. Since many of those who relapse eventually return to abstinence (e.g. Pickens *et al.*, 1985), it has been suggested that relapses may serve a useful function by providing opportunities to learn how to handle high risk situations (Brownell *et al.*, 1986). This interesting possibility requires further study.

Clearly, following-up on any of these ideas will require a much better understanding of the relapse process itself which, in turn, as we have learned from the studies in this issue, will require better instrumentation and improved methodology to conduct the necessary research. With respect to the question posed in my title, the Marlatt coding system as originally proposed has not been demonstrated to be either reliable or valid by the current series of studies. However, it seems to me that reliability and validity of some of the categories of Marlatt's taxonomy, broadly defined (e.g. "negative emotions"), have been supported to at least some extent by a number of independent studies, including a few of those reported in this issue. These independent findings suggest replicability across studies and construct validity of some of the major categories originally identified by Marlatt as being the most important ones.

Given these fairly consistent strands in tests of Marlatt's cognitive–behavioral conceptualization of the relapse process, it appears likely that the general framework of his taxonomy (along the lines of the Level 2 categories) will ultimately be supported. Nevertheless, some of the categories will need to be modified in light of accumulating evidence, and some of the alternative assessment instruments that have been developed may prove superior for research purposes to Marlatt's original coding scheme. From the clinical perspective, however, there may be a trade-off. As greater psychometric precision is gained through the use of more structured instruments, there may be a loss in richness of detail and of clients' attributions for their relapses, information that is more likely to be obtained with Marlatt's open-ended questions. As the refinement process proceeds, these sometimes conflicting clinical and research needs will have to be accommodated.

Acknowledgements

This commentary derives from invited discussant remarks that were initially presented as part of a symposium at the Seventh International Conference on Treatment of Addictive Behaviors, The Netherlands, May 1995. Supported in part by NIAAA grant RO1-AA09648.

References

ANNIS, H. M. (1982) *Inventory of Drinking Situations (IDS-100)* (Toronto, Ontario, Addiction Research Foundation of Ontario).

ANNIS, H. M. & DAVIS, C. S. (1988) Self-efficacy and the prevention of alcoholic relapse: initial findings from a treatment trial, in: BAKER, T. B. & CANNON, D. S. (Eds) *Assessment and Treatment of Addictive Disorders*, pp 88–112 (New York, Praeger Press).

ANNIS, H. M. & DAVIS, C. S. (1989) Relapse prevention, in: HESTER, R. K. & MILLER, W. R. (Eds) *Handbook of Alcoholism Treatment Approaches*, pp. 170–182 (New York, Pergamon Press).

ANNIS, H. M. & GRAHAM, J. M. (1995) Profile types on the inventory of drinking situations: implications for relapse prevention counseling, *Psychology of Addictive Behaviors*, 9, 176–182.

BAER, J. S. & LICHTENSTEIN, E. (1988) Classification and prediction of smoking relapse episodes: an exploration of individual differences, *Journal of Consulting and Clinical Psychology*, 56, 104–110.

BROWNELL, K. D., MARLATT, G. A., LICHTENSTEIN, E. & WILSON, G. T. (1986) Understanding and preventing relapse, *American Psychologist*, 41, 765–782.

CANNON, D. S., LEEKA, J. K., PATTERSON, E. T. & BAKER, T. B. (1990) Principal components analysis of the Inventory of Drinking Situations: empirical categories of drinking by alcoholics, *Addictive Behaviors*, 15, 265–269.

CHANEY, E. F., O'LEARY, M. R. & MARLATT, G. A. (1978) Skill training with alcoholics, *Journal of Consulting and Clinical Psychology*, 46, 1092–1104.

HEATHER, N. & STALLARD, A. (1989) Does the Marlatt model underestimate the importance of conditioned craving in the relapse process? *Relapse and Addictive Behaviour*, pp. 180–208 (New York, Tavistock/Routledge).

HEATHER, N., STALLARD, A. & TEBBUTT, J. (1991) Importance of substance cues in relapse among heroin users: comparison of two methods of investigation, *Addictive Behaviors*, 16, 41–49.

HODGINS, D. C., EL-GUEBALY, N. & ARMSTRONG, S. (1995) Prospective and retrospective reports of mood states before relapse to substance use, *Journal of Consulting and Clinical Psychology*, 63, 400–407.

ISENHART, C. E. (1991) Factor structure of the Inventory of Drinking Situations, *Journal of Substance Abuse*, 3, 59–71.

LITMAN, G. K., STAPLETON, J., OPPENHEIM, A. N., PELEG, M. & JACKSON, P. (1983) Situations related to alcoholism relapse, *British Journal of Addiction*, 78, 381–389.

LONGABAUGH, R., RUBIN, A., STOUT, R. L., ZYWIAK, W. H. & LOWMAN, C. (1996) The reliability of Marlatt's taxonomy for classifying relapses, *Addiction*, 91: Supplement, S73–S88.

MAISTO, S. A., CONNORS, G. J. & ZYWIAK, W. H.

(1996) Construct validation analyses on the Marlatt typology of relapse precipitants, *Addiction*, 91: Supplement, S89–S98.

MARLATT, G. A. (1996) Taxonomy of high-risk situations for alcohol relapse: evolution and development of a cognitive-behavioural model of relapse, *Addiction*, 91: Supplement, S37–S50.

MARLATT, G. A. & GORDON, J. R. (1980) Determinants of relapse: implications for the maintenance of behavior change, in: DAVIDSON, P. & DAVIDSON, S. M. (Eds) *Behavioral Medicine: Changing Health Lifestyles*, pp. 410–452 (New York, Brunner/Mazel).

PICKENS, R. W., HATSUKAMI, D. K., SPICER, J. W. & SVIKIS, D. S. (1985) Relapse by alcohol abusers, *Alcoholism: Clinical and Experimental Research*, 9, 244–247.

SANDAHL, C. (1984) Determinants of relapse among alcoholics: a cross-cultural replication study, *International Journal of the Addictions*, 19, 833–848.

STOUT, R. L., LONGABAUGH, R. & RUBIN, A. (1996) Predictive validity of Marlatt's taxonomy versus a more general relapse code, *Addiction*, 91: Supplement, S99–S110.

ZYWIAK, W. H., CONNORS, G. J., MAISTO, S. A. & WESTERBERG, V. S. (1996) Relapse research and the Reasons for Drinking Questionnaire: a factor analysis of Marlatt's taxonomy, *Addiction*, 91: Supplement, S121–S130.