

Application of Machine Learning to Predict Dietary Lapses During Weight Loss

Journal of Diabetes Science and Technology
2018, Vol. 12(5) 1045–1052
© 2018 Diabetes Technology Society
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1932296818775757
journals.sagepub.com/home/dst



Stephanie P. Goldstein, MS¹, Fengqing Zhang, PhD²,
John G. Thomas, PhD³, Meghan L. Butryn, PhD¹,
James D. Herbert, PhD⁴, and Evan M. Forman, PhD¹

Abstract

Background: Individuals who adhere to dietary guidelines provided during weight loss interventions tend to be more successful with weight control. Any deviation from dietary guidelines can be referred to as a “lapse.” There is a growing body of research showing that lapses are predictable using a variety of physiological, environmental, and psychological indicators. With recent technological advancements, it may be possible to assess these triggers and predict dietary lapses in real time. The current study sought to use machine learning techniques to predict lapses and evaluate the utility of combining both group- and individual-level data to enhance lapse prediction.

Methods: The current study trained and tested a machine learning algorithm capable of predicting dietary lapses from a behavioral weight loss program among adults with overweight/obesity ($n = 12$). Participants were asked to follow a weight control diet for 6 weeks and complete ecological momentary assessment (EMA; repeated brief surveys delivered via smartphone) regarding dietary lapses and relevant triggers.

Results: WEKA decision trees were used to predict lapses with an accuracy of 0.72 for the group of participants. However, generalization of the group algorithm to each individual was poor, and as such, group- and individual-level data were combined to improve prediction. The findings suggest that 4 weeks of individual data collection is recommended to attain optimal model performance.

Conclusions: The predictive algorithm could be utilized to provide in-the-moment interventions to prevent dietary lapses and therefore enhance weight losses. Furthermore, methods in the current study could be translated to other types of health behavior lapses.

Keywords

obesity, diet, ecological momentary assessment, machine learning

Within behavioral weight loss programs, dietary lapses (ie, deviations from dietary recommendations) are common and associated with poor outcomes.^{1,2} Ecological momentary assessment (EMA) has been used to identify psychological, physiological, and environmental factors associated with dietary lapses through contextually valid assessment.¹⁻⁴ EMA studies can inform just-in-time adaptive interventions (JITAIs) for dietary lapses that assist individuals in modifying environmental cues and coping with triggers directly in moments of need.^{5,6} However, further research is needed to evaluate the methods for real-time lapse predictions.

Real-time prediction can be achieved via machine learning, which involves the development of computer systems that can adapt based on incoming data. One study to date has utilized machine learning to predict overeating episodes with

71.3% accuracy based on self-reported data of 16 non-treatment-seeking adults with overweight/obesity.⁷ The current

¹Center for Weight, Eating, and Lifestyle Science and Department of Psychology, Drexel University, Philadelphia, PA, USA

²Department of Psychology, College of Arts and Sciences, Drexel University, Philadelphia, PA, USA

³Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, The Miriam Hospital Weight Control and Diabetes Research Center, Providence, RI, USA

⁴President's Office, University of New England, Biddeford, ME, USA

Corresponding Author:

Stephanie P. Goldstein, MS, Center for Weight, Eating, and Lifestyle Science and Department of Psychology, Drexel University, 3141 Chestnut St, Stratton 119, Philadelphia, PA 19104, USA.
Email: spg38@drexel.edu

project sought to replicate and extend this prior work by evaluating an algorithm specifically targeting lapses from a weight loss diet among a treatment-seeking sample. Our methods are novel in that we investigate the utility of combining group- and individual-level data to efficiently generate real-time lapse predictions. This procedure can reduce the time spent on data collection while preserving the ability to detect patterns in the data.

We recruited a small sample ($n = 12$) of weight-loss-seeking individuals with overweight or obesity. Participants were asked to adhere to the Weight Watchers® (WW) weight loss plan and continuously self-report lapses and relevant triggers through an EMA smartphone app for 6 weeks. It was estimated that 6 weeks of data collection from 12 individuals would provide adequate data for a preliminary model (an estimated 216-288 lapses based on prior work)⁸ while also piloting our methodology for algorithm development.

Study Aims

Aim 1

Our primary aim was to achieve a model for predicting dietary lapses that achieves a priori thresholds of accuracy, sensitivity, and specificity. An accuracy threshold was set at $>70\%$ based on extant literature.^{9,10} The threshold for sensitivity (ie, accurate prediction of lapse) was set at $>70\%$ so that the algorithm could predict the majority of lapses. Strong sensitivity typically comes at the expense of specificity (eg, accurate prediction of a non-lapse).¹¹ Therefore, we set a threshold of $>50\%$ specificity to minimize false positives.

Aim 2

The next aim was to evaluate the ability of separate individual-level models to predict dietary lapses. We also tested the ability of a group model to predict lapses for a completely new user. Given the individualized nature of eating behavior and the rarity of lapse cases, we hypothesized that neither group- nor individual-level models alone will be capable of meaningful prediction.

Aim 3

The final aim sought to evaluate a combination of group- and individual- level data to enhance prediction.

Methods

Participants

This study included 12 participants (age mean = 38.25, SD = 13.54; 91.7% female; 50.0% white, 50.0% black/ African American) with overweight or obesity (BMI mean = 33.60 kg/m², SD = 5.66). Inclusion criteria were age 18-65, body mass index 27-45 kg/m², and ownership of an iOS device with data plan.

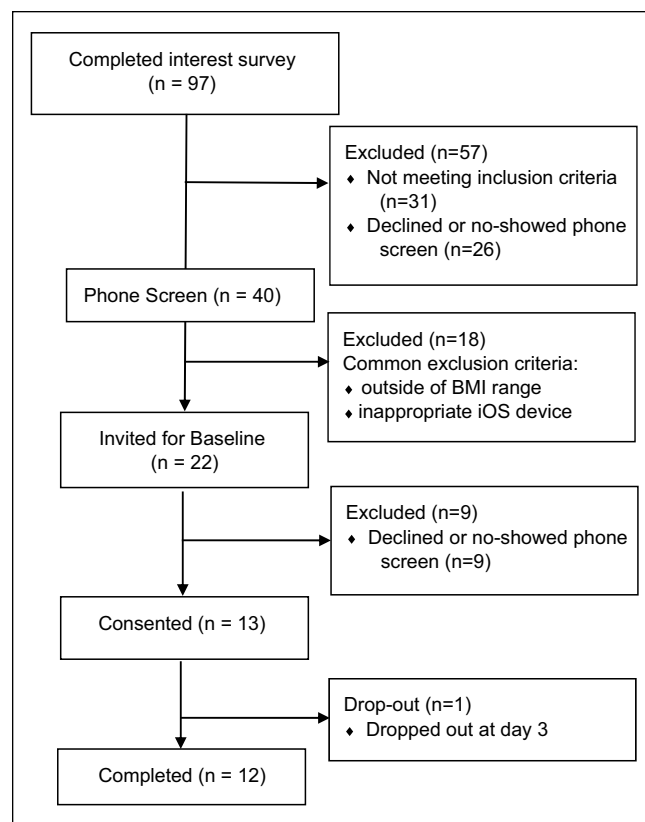


Figure 1. CONSORT diagram of study flow.

Exclusion criteria were participation in a structured weight loss program, pregnancy, serious medical conditions that affected weight or appetite, reported disordered eating symptoms, or medication change known to affect weight or appetite within the last 3 months.

Procedures

See Figure 1 for depiction of study flow. Participants were recruited from Philadelphia and surrounding areas using list-servs and community postings. Interested participants were screened using an online survey. Additional screening occurred via telephone. Eligible individuals were invited to attend an in-person baseline appointment. At this appointment, informed consent was obtained from all participants and documented according to the specifications of the US Department of Health and Human Services and US Food and Drug Administration.

After providing informed consent, participants were given the EMA smartphone app and the WW program at no cost and were taught to use both. The WW program, an evidence-based weight loss program,¹² assigns points goals per day and these daily point goals were further broken down into specific meal/snack targets. Participants were instructed to report a dietary lapse in the EMA smartphone app if they exceeded any of their meal/snack point targets throughout the day. This procedure was developed to enhance the validity of a lapse

Table 1. Lapse Predictor EMA Measures.

Predictor	Question	Answer scale
Affect	Rate your current mood:	0 = much better than my usual self, 4 = much worse than my usual self
Boredom	How bored do you feel right now?	0 = not at all, 4 = extremely
Hunger	How hungry are you right now?	0 = not at all, 4 = extremely
Fatigue	How tired are you right now?	0 = not at all, 4 = extremely
Hours of sleep	How many hours of sleep did you have last night?	0 = <4 hours, 7 = >8 hours
Cravings	If you have experienced a craving for a specific food, how strong was the craving?	0 = not strong at all, 4 = extremely strong
Urges	Since the last prompt, have you had a sudden urge to go off your eating plan for the day (regardless of whether you acted on this or not)?	Yes/no
Cognitive load	In the last hour, think of the most difficult task you were working on. How difficult was this task in terms of the mental effort required (eg, figuring, planning, decision-making)?	0 = requiring almost no mental effort, 4 = requiring almost all of my mental effort
Confidence	How confident are you that you can meet your dietary goals for the rest of the day?	0 = not at all, 6 = extremely
Motivation	How motivated are you to follow your dietary plan right now?	0 = not at all, 1 = somewhat, 2 = very
Socializing	In the past hour, have you engaged in socializing with coworkers, family, or friends?	0 = no, 1 = yes, without food, 2 = yes, with food
TV	Are you watching TV right now?	Yes/no
Exercise	Have you engaged in structured exercise (eg, setting aside time specifically devoted to exercising) today?	Yes/no
Neg interpersonal interactions	In the past hour, have you had an unpleasant encounter with another person?	Yes/no
Tempting food availability	Over the past hour, has there been tempting food/drink within close reach?	Yes/no
Alcohol	Have you consumed any alcohol today?	Yes/no
Food ads	In the past hour, have you seen an advertisement for food?	Yes/no
Planning	To what extent have you planned exactly what you will eat for rest of the day?	0 = not at all, 1 = somewhat, 2 = very
Missed meal/snack	Since the last survey, have you eaten a meal/snack?	0 = no, 1 = yes, 2 = unsure

report by increasing objectivity. Participants were retrained on the definition of a dietary lapse at a follow-up appointment 3-5 days after baseline.

Participants used the EMA smartphone app and the WW program for 6 weeks. They received weekly emails with information about survey completion, study earnings, and WW mobile app use. Participants also received 20-minute phone calls in weeks 2 and 4 to reinforce good compliance with EMA and troubleshoot problems with the apps. Participants received payment for participation every other week. Consistent with prior EMA studies, participants could earn up to \$180 total, with \$0.50 deduction for every missed EMA prompt.¹³⁻¹⁶ Participants' monetary balances were displayed in the EMA app.

Ecological Momentary Assessment

The EMA app prompted participants at 6 quasi-random times to enter data (time-based sampling) and allowed participants to give unprompted lapse reports (event-based sampling). Over 20 potential lapse triggers were identified from the literature and assessed. Asking participants to respond to 20

questions 6 times per day was judged infeasible. Instead, each prompt contained 8 questions drawn quasi-randomly. The frequency and timing of questions varied based on theory and research support (eg, hours of sleep was asked once in the morning, mood was asked several times per day as it tends to vary often).¹⁷ If participants responded "yes" to a lapse, a subset of questions were replaced with specific lapse-related questions (eg, time, location) so as not to penalize lapse reporting. Surveys were timestamped upon completion.

Measures

Dietary Lapse. Dietary lapses were defined as any instance in which an individual exceeded a meal or snack point target. Participants also recorded the time and date of each lapse.

Lapse Predictors. Lapse predictors were identified via a literature search of factors previously associated with lapses, general overeating behavior, or poor self-control. Questions were developed by the study team based on our prior EMA work.⁸ See Table 1 for a complete list of predictors, questions, and answer scales.

Trait-Like Predictive Factors. A number of trait-like factors can impact eating behavior and therefore were included in our models. A demographics questionnaire, given at baseline, included questions on age, sex, ethnicity. Weight and height was measured using a research-grade calibrated scale and stadiometer. Dieting and weight history was assessed via self-report of current and previous dieting attempts, and difference between current and lowest weight (ie, weight suppression). Restraint and overeating were assessed using the Eating Inventory (EI),¹⁸ formally the Three-Factor Eating Questionnaire. The EI contains 51 items and is composed of three subscales including cognitive restraint (ie, an individual's presumed cognitive control over eating behavior), disinhibition (ie, eating in excess in response to emotional and cognitive cues), and hunger. EI factors were included in the model because they are associated with real-world eating behavior.¹⁹ Responsivity to food was measured using the Power of Food Scale (PFS).²⁰ The PFS is a 15-item self-report measure that assesses individual differences in the psychological influence of the food environment. The PFS predicts cravings and consumption among individuals who are dieting.²¹ Propensity for food cravings was measured using the Food Cravings Questionnaire–Trait (FCQ-T).²² This 39-item self-report measure prompts respondents to indicate how often each statement regarding food cravings is true for them on a 6-point scale (from *never* to *always*). Strength and frequency of food cravings tend to predict consumption (eg, susceptibility to dietary lapse).²¹

Statistical Approach

R version 3.1.2 was used to analyze data.

Preparation of Data. Predictors were lagged onto the subsequent case to reflect event prediction. Lagging variables from one day to the next was not performed. The data set was imbalanced with non-lapses outnumbering lapse instances by 1:9.5. To prevent the algorithm from focusing on the most numerous case (ie, non-lapse) rather than the lapses, the ROSE package was used to balance the data.^{23,24}

Our EMA procedure to alleviate participant burden resulted in systematic missing data (ie, a response was not possible because of quasi-random question administration) in addition to the data missing due to nonresponse. A separate class approach to imputation was used, which is recommended when there is substantial missing data in both the training and testing data sets.²⁵ Using this method, categorical and continuous variables received codes specifying missing data.

Training and Testing. Training and testing of data was conducted through cross-validation.²³ The model was trained on the first 4 weeks of data and made a prediction on the subsequent 2 weeks of data to simulate how predictions would be used when the algorithms are ultimately applied in the context of

treatment. The goal was to provide the algorithm with as much training data as possible while preserving adequate data for testing. Predicted outcomes were compared to the actual outcomes to estimate accuracy, sensitivity, and specificity. Estimates were compared to a priori thresholds for acceptable sensitivity (.70), accuracy (.70), and specificity (.50).

Model performance was also evaluated by calculating an area under the curve (AUC) estimate using the pROC package²⁶ to conduct a receiver operating characteristic (ROC) analysis. An AUC of 1 indicates a perfect test, while .50 represents a test that is no better than chance.^{11,27} Furthermore, 95% confidence intervals (CIs) of AUC values were calculated to estimate stability. If the CI included .50, model estimates were deemed unreliable.²⁸

Aim 1: Model Selection. The optimal group model was identified using ensemble methods (eg, combining weighted vote of predictions from Random Forest, Logit.Boost, Bagging, Random Subspace, Bayes Net). Cost-sensitive methods were used by incorporating a cost matrix (eg, a matrix of penalties for misclassification) into each decision tree.²⁹ Cost sensitive penalties were selected based on a balance of sensitivity and specificity (eg, highest possible sensitivity while maintaining adequate specificity).

Aim 2: Individual- and Group-Level Analyses. Accuracy, sensitivity, and specificity of the classification model built within each participant was examined. To examine model generalization, “leave-one-out” cross-validation procedures were used in which an entire participant's set of responses was left out of the training set and used as the testing set. This analysis assessed the ability of a model built on the group-level (training set including multiple subjects) to predict a new individual's observations. The number of models that met threshold criteria and AUC values were calculated for individual and group cross-validation.

Aim 3: Combining Group- and Individual-Level Data. The above analyses were repeated with cross-validation that included part of the data from the testing case in the training data. Therefore, the model was trained using data from the group, as well as part of the data from the testing case. The final cross-validation step was tested with 4 weeks of individual data added to the group. The number of models that met threshold criteria and AUC values were calculated.

Results

Descriptive Information

See Table 2 for descriptive information of model predictors. Collectively, participants reported a total of 292 lapses and there were 2,551 non-lapse ratings. Participants reported an average of 3.47 lapses per week (SD = 2.41, 95% CI [2.10, 4.84]). An average of 24.33 lapses per

Table 2. Descriptive Information From Model Predictors.

EMA variable (total responses)	Scale	Mean	SD	n	%
Hours of sleep (n = 481)	0-7	5.3	2.12		
Boredom (n = 1208)	0-4	0.83	1.17		
Hunger (n = 1282)	0-4	1.45	1.41		
Cravings (n = 1290)	0-4	0.97	1.22		
Fatigue (n = 1280)	0-4	1.84	1.48		
Urges (n = 1238)				209	16.9
Confidence (n = 478)	0-6	3.29	1.37		
Motivation (n = 1452)	0-2	1.50	0.87		
Cognitive load (n = 992)	0-4	1.93	1.12		
Planning (n = 473)	0-2	1.90	1.06		
Affect (n = 1518)	0-4	1.77	0.68		
Tempting food availability (n = 1607)				564	35.1
Socializing without food present (n = 946)				230	24.3
Socializing with food present (n = 946)				319	33.7
TV (n = 940)				357	38.0
Neg interpersonal interactions (n = 979)				64	6.5
Food ads (n = 953)				163	17.1
Healthy food availability (n = 982)				611	62.2
Exercise (n = 471)				162	34.3
Alcohol (n = 473)				32	6.8
Missed meal/snack (n = 2564)				1252	48.8
Unsure if missed meal/snack (n = 2564)				16	0.01
Moderator variables		Mean	SD	n	
Weight suppressed 1-2 lbs.				2	16.7
Weight suppressed 5-6 lbs.				1	8.3
Weight suppressed 7-8 lbs.				1	8.3
Weight suppressed > 10lbs.				8	66.7
Dieting at baseline				0	0
Past dieting				7	58.3
External disinhibition (EI)	0-6	5.50	1.62		
Internal disinhibition (EI)	0-8	2.0	1.47		
Restraint (EI)	0-21	16.40	6.03		
Power of food	1-5	2.30	0.63		
Food cravings-trait	1-6	2.50	0.75		

person were reported (SD = 15.82, range = 0-48). Participants responded to an average of 94.6% of EMA prompts (range = 85.2%-98.9%) and compliance remained relatively stable throughout the study.

Aim 1

The final group model accuracy (0.72) exceeded the a priori threshold while maintaining good sensitivity (0.70) and specificity (0.72). In addition, the ROC analysis (AUC = 0.72, 95% CI_{AUC} [0.66, 0.77]) indicated that the group-level model was adequate.

Aim 2

To accomplish Aim 2, the performance of group-level and individual-level models were examined separately. Model

outcomes from each analysis are depicted in Figure 2. One participant (participant A) was removed from the below analyses because no lapses were reported for the duration of the study.

Individual-Level Model. To examine the effect of an individual-level model, a model was built based on each participant's first 4 weeks of data (and tested on the remaining two). This procedure resulted in 11 sets of model outcomes (one for each participant who had reported enough lapse cases). Two individual-level models met the minimum criteria for model adequacy. Furthermore, 95% CIs of the AUC values included 0.50 (indicating model instability) for all but one participant.

Group-Level Model. To examine the effect of the group-level model, one participant was left out of model construction and the algorithm was tested on that particular individual.

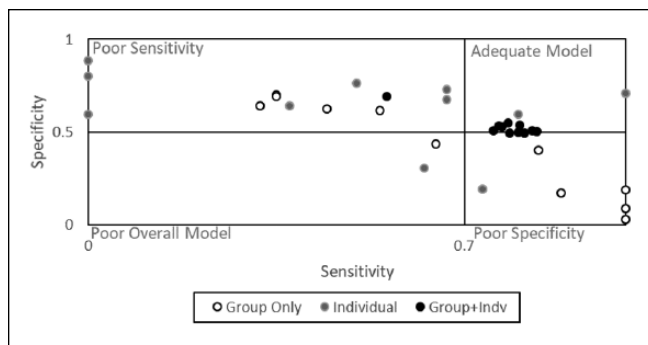


Figure 2. Model outcomes summarized across group, individual and group + individual data.

This procedure resulted in 11 sets of model outcomes (one for each participant who had reported enough lapse cases). None of the group-level models tested on each individual participant met the minimum criteria for model adequacy (specificity $\geq .50$ and sensitivity $\geq .70$). An examination of 95% CIs of the AUC values for these models indicated that approximately half of models (54.5%) had CIs including 0.50 (indicating model instability), whereas the remaining 45.5% had more stable CIs (eg, did not include .50).

Aim 3

To further examine the effect of the combined group- and individual-level model (Aim 3), each participant's data was again split into the first 4 weeks and the last 2 weeks. Model construction involved the group's data in addition to the first 4 weeks of an individual's responses and the algorithm was tested on that particular individual's final 2 weeks of responses. This procedure resulted in 11 sets of model outcomes (one for each participant who had reported enough lapse cases). Eight models (constructed from the group data with 4 weeks of an individual's data) met the minimum criteria for model adequacy. Furthermore, 95% CIs of the AUC excluded 0.50 for all participants, indicating good model stability. Model outcomes are depicted in Figure 2.

Discussion

The current study sought to develop a classification algorithm to predict dietary lapses among individuals who were following a weight loss diet. Our data collection methods (eg, length of time, ease of reporting) were tolerable to participants as evidenced by good compliance throughout the study. The average number of lapse reports was consistent with prior literature.¹⁻⁴ Approximately 300 lapse cases were collected, which surpassed the minimum number of 75 cases recommended for classification.³⁰

The algorithm that resulted from the current study can predict when an individual is likely to lapse from his or her weight control diet over the next several hours based on self-reported

factors. These results met the minimum criteria specified for a clinically relevant algorithm that is better than chance. Secondly, our study aimed to investigate the predictive ability of both group- and individual-level data as combining the two data types is a commonly cited advantage of machine learning.³¹ Results illustrated that the group-level models (eg, predicting one participant's behavior using an entirely separate group's data) generally had either poor sensitivity or poor specificity. For individual-level models, CIs were generally unstable and models suffered from poor sensitivity. Of note, model outcomes showed the greatest improvement when 4 weeks of individual data was added to the group-level data. Such results have strong implications for the translation of this algorithm into a JITAI for lapses. For instance, results indicate that the ideal JITAI for lapses would have a combination of group- and individual-level data. This could involve a temporary period in which an individual provides data (without intervention) until enough data are accrued to make adequate predictions when combined with the group.

Strengths and Limitations

The current study is novel in that it is the first to apply machine learning techniques to predict of dietary lapses. In addition, the study is innovative in that it specifically sought to create a scalable set of algorithms with clinical utility. The methods used here can apply broadly to JITAI solutions for other weight-related conditions (eg, diabetes, binge eating) given the similarities of these behaviors to dietary lapse and parallels in form of intervention.

The present study also contained several limitations. First, our data collection procedure may have impaired model performance by creating large amounts of missing data that were difficult to impute.^{32,33} Second, the small sample size raises questions regarding the generalizability of findings to the populations with overweight/obesity. CIs and standard deviations reveal that our outcomes are stable and reliable. Despite this reliability, results do indicate that group models generalize poorly to each individual. Thus, it appears that our sample generated enough data to create a successful machine learning model, but (as would be predicted) they are not representative of the population. It is unknown if more, and how many more, participants would be required to achieve a model that has better generalizability, and this should be an area for future study. Last, research on the prediction of lapses within other types of weight loss dieting (besides Weight Watchers) is necessary given that this algorithm and approach may not generalize to all diet programs and it may be more difficult for participants to conceptualize and report lapses in other programs.

Conclusion

In sum, machine learning techniques appear to be a promising solution for the real-time prediction of dietary lapse. The

current study is a proof of concept for the application of machine learning techniques to the problem of lapsing. Continued research is necessary to refine lapse prediction by developing a more finite set of lapse predictors or utilizing predictors that require less input from the participant.³⁴⁻³⁷

The next step of this research is to use this algorithm to power a JITAI that preventatively delivers interventions for lapses based on continuous model predictions. One method of testing the efficacy of this JITAI would be a randomized controlled trial comparing the JITAI to an EMA-only condition to control for reactivity to repeated assessments. A more stringent comparison condition would be provision of non-tailored interventions. Trials such as these should be powered to detect between-groups differences in weight loss as mediated by objectively measured dietary adherence. With continued innovations in the fields of mobile phone technology and data mining, JITAIs can become powerful tools for health behavior change.

Abbreviations

AUC, area under the curve; CI, confidence interval; EI, Eating Inventory; EMA, ecological momentary assessment; FCQ-T, Food Cravings Questionnaire–Trait; JITAI, just-in-time adaptive intervention; PFS, Power of Food Scale; ROC, receiver operating characteristic; WW, Weight Watchers.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: APAGS/Psi Chi Junior Scientist Fellowship awarded to Stephanie Goldstein. The Obesity Society/Weight Watchers Karen Miller-Kovach Research Grant awarded to Evan Forman.

References

- Forman EM, Schumacher LM, Crosby R, et al. Ecological momentary assessment of dietary lapses across behavioral weight loss treatment: characteristics, predictors, and relationships with weight change. *Ann Behav Med*. 2017;51(5):1-13.
- McKee HC, Ntoumanis N, Taylor IM. An ecological momentary assessment of lapse occurrences in dieters. *Ann Behav Med*. 2014;48(3):300-310.
- Carels RA, Douglass OM, Cacciopaglia HM, O'Brien WH. An ecological momentary assessment of relapse crises in dieting. *J Consult Clin Psychol*. 2004;72(2):341-348.
- Carels RA, Hoffman J, Collins A, Raber AC, Cacciopaglia H, O'Brien WH. Ecological momentary assessment of temptation and lapse in dieting. *Eat Behav*. 2002;2(4):307-321.
- Hofmann W, Dohle S. Capturing eating behavior where the action takes place: a comment on McKee et al. *Ann Behav Med*. 2014;48(3):289-290.
- Spuijdt-Metz D, Nilsen W. Dynamic models of behavior for just-in-time adaptive interventions. *IEEE Pervasive Comput*. 2014;13(3):13-17.
- Tulu B, Ruiz C, Allard J, et al. SlipBuddy: a mobile health intervention to prevent overeating. In: *Proceedings of the 50th Hawaii International Conference on System Sciences*. 2017. Retrieved from <http://dx.doi.org/10.24251/HICSS.2017.436>
- Forman EM, Schumacher LM, Crosby R, et al. Ecological momentary assessment of dietary lapses across behavioral weight loss treatment: characteristics, predictors, and relationships with weight change. *Ann Behav Med*. 2017;51(5):741-753.
- Swets J. *Evaluation of Diagnostic Systems*. Dordrecht, Netherlands: Elsevier; 2012.
- Fisher SRA, Fisher RA, Genetiker S, et al. *The Design of Experiments*. Edinburgh, UK: Oliver and Boyd; 1960.
- Fan J, Upadhye S, Worster A. Understanding receiver operating characteristic (ROC) curves. *CJEM*. 2006;8(1):19-20.
- Dansinger ML, Gleason JA, Griffith JL, Selker HP, Schaefer EJ. Comparison of the Atkins, Ornish, Weight Watchers, and Zone diets for weight loss and heart disease risk reduction: a randomized trial. *JAMA*. 2005;293(1):43-53.
- Jeffers AJ. *Ecological Momentary Assessment and Time-Varying Factors Associated With Eating and Physical Activity* [master's thesis]. Richmond: Virginia Commonwealth University; 2012.
- Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol*. 2008;4:1-32.
- Stone AA, Shiffman S. Ecological momentary assessment (EMA) in behavioral medicine. *Ann Behav Med*. 1994;16:199-202.
- Wong JHK. *Capturing the Dynamics of a Work Day: Ecological Momentary Assessment of Work Stressors on the Health of Long-Term Caregivers*. Ottawa: Library and Archives Canada; 2012.
- Goldstein SP, Evans BC, Flack D, Juarascio A, Manasse S, Zhang F, et al. Return of the JITAI: applying a just-in-time adaptive intervention framework to the development of m-health solutions for addictive behaviors. *Int J Behav Med*. 2017;24(5):673-682.
- Stunkard AJ, Messick S. *Eating Inventory: Manual*. New York, NY: Psychological Corporation; 1988.
- Clark MM, Marcus BH, Pera V, Niaura RS. Changes in eating inventory scores following obesity treatment. *Int J Eat Disord*. 1994;15(4):401-405.
- Lowe MR, Butryn ML, Didie ER, et al. The Power of Food Scale. A new measure of the psychological influence of the food environment. *Appetite*. 2009;53(1):114-118.
- Forman EM, Hoffman KL, McGrath KB, Herbert JD, Brandsma LL, Lowe MR. A comparison of acceptance-and control-based strategies for coping with food cravings: an analog study. *Behav Res Ther*. 2007;45(10):2372-2386.
- Cepeda-Benito A, Gleaves DH, Williams TL, Erath SA. The development and validation of the state and trait food-cravings questionnaires. *Behav Ther*. 2001;31(1):151-173.
- Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage*. 2009;45(1):S199-S209.
- Lunardon N, Menardi G, Torelli N. *ROSE: A Package for Binary Imbalanced Learning*. Vienna, Austria: R Foundation for Statistical Computing; 2015.

25. Ding Y, Simonoff JS. An investigation of missing data methods for classification trees applied to binary response data. *J Mach Learn Res*. 2010;11:131-170.
26. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 2011;12:77.
27. Mohri C. Confidence intervals for the area under the ROC curve. In: *NIPS'04 Proceedings of the 17th International Conference on Neural Information Processing Systems*. Cambridge, MA: MIT Press; 2005:305-312.
28. Fischer JE, Bachmann LM, Jaeschke R. A readers' guide to the interpretation of diagnostic test properties: clinical example of sepsis. *Intensive Care Med*. 2003;29(7):1043-1051.
29. Ling CX, Sheng VS. Cost-sensitive learning and the class imbalance problem. In: Sammut C, ed. *Encyclopedia of Machine Learning*. New York, NY: Springer; 2011:231-235.
30. Beleites C, Neugebauer U, Bocklitz T, Krafft C, Popp J. Sample size planning for classification models. *Anal Chim Acta*. 2013;760:25-33.
31. Witten IH, Frank E. *Data mining: practical machine learning tools and techniques*. San Francisco, CA: Morgan Kaufmann; 2005.
32. Lee JH, Huber J Jr. *Multiple imputation with large proportions of missing data: How much is too much?* United Kingdom Stata Users' Group Meetings 2011. College Station: Texas A&M Health Science Center; 2011.
33. Ziegler ML. *Variable Selection When Confronted With Missing Data*. Pittsburgh, PA: University of Pittsburgh; 2006.
34. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. *Psychiatr Rehabil J*. 2015;38(3):218-226.
35. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as an electronic biomarker of illness activity in bipolar disorder. *Bipolar Disord*. 2015;17(7):715-728.
36. Luxton DD, June JD, Sano A, Bickmore T. Intelligent mobile, wearable, and ambient technologies for behavioral health care. In: Luxton DD, ed., *Artificial Intelligence in Behavioral and Mental Health Care*. San Diego, CA: Elsevier Academic Press; 2015:137.
37. Gravenhorst F, Muaremi A, Bardram J, et al. Mobile phones as medical devices in mental disorder treatment: an overview. *Pers Ubiquitous Comput*. 2015;19(2):335-353.