Predictive modeling of addiction lapses in a mobile health application ☆

Ming-Yuan Chih, M.H.A., M.S. *, Timothy Patton, B.S., Fiona M. McTavish, M.S., Andrew J. Isham, M.S., Chris L. Judkins-Fisher, B.S., Amy K. Atwood, Ph.D., David H. Gustafson, Ph.D.

Center for Health Enhancement Systems Studies, College of Engineering, University of Wisconsin – Madison, 1513 University Ave., Madison, WI 53706, USA

ABSTRACT

The chronically relapsing nature of alcoholism leads to substantial personal, family, and societal costs. Addiction-comprehensive health enhancement support system (A-CHESS) is a smartphone application that aims to reduce relapse. To offer targeted support to patients who are at risk of lapses within the coming week, a Bayesian network model to predict such events was constructed using responses on 2,934 weekly surveys (called the Weekly Check-in) from 152 alcohol-dependent individuals who recently completed residential treatment. The Weekly Check-in is a self-monitoring service, provided in A-CHESS, to track patients' recovery progress. The model showed good predictability, with the area under receiver operating characteristic curve of 0.829 in the 10-fold cross-validation and 0.912 in the external validation. The sensitivity/specificity table assists the tradeoff decisions necessary to apply the model in practice. This study moves us closer to the goal of providing lapse prediction so that patients might receive more targeted and timely support.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

In 2010, an estimated 22.1 million people (8.7% of the U.S. population aged 12 or older) in the U.S. were classified with alcohol dependence or abuse (Substance Abuse and Mental Health Services Administration, 2011). Patients striving to change problematic drinking behavior often experience an initial setback, a transitory state or so-called lapse, that can lead to a full return to their previous drinking behavior (known as a relapse) (Witkiewitz & Marlatt, 2004). Relapse reduces the quality of life for the drinking individuals and their families, and can place a huge burden on society through crime, accident, healthcare costs, and reduced productivity (Sullivan, Fiellin & O'Connor, 2005). To prevent relapse, researchers have called for more adaptive interventions that offer constant monitoring, a wide variety of services, and decision support for clinicians to help them adjust the services to meet the changing needs of patients (McKay, 2005, 2009).

To address this issue, the Center for Health Enhancement Systems Studies (CHESS) at the University of Wisconsin-Madison developed a smartphone-based, relapse-prevention program, addiction-comprehensive health enhancement support system (A-CHESS) (Gustafson, Boyle, et al., 2011). Various services were designed and included in A-CHESS to promote patients' autonomous motivation, coping competence, and relatedness, so that patients were prepared to face different challenges along their recovery journey (Gustafson, Shaw, et al., 2011; Marlatt & George, 1984; Ryan & Deci, 2000). For example, patients can connect with their peers and counselors at any time via these A-CHESS services: "discussions", "my messages", "team feed", and "support team". Patients can search the latest addiction-related news articles and access other credible Web resources in addiction recovery in "news" and "recovery info"; listen to the stories of other patients from "our stories" and "recovery podcasts"; find the nearest alcoholics (or narcotics) anonymous meetings from "meetings"; use "panic button" and "easing distress" to help them cope with their recovery; and track their progress in "surveys". More details on A-CHESS services can be found in other publications (Gustafson, Boyle, et al., 2011; Gustafson, Shaw, et al., 2011; McTavish, Chih, Shah, & Gustafson, 2012).

A randomized trial to test the efficacy of A-CHESS on patients with a diagnosis of DSM-IV alcohol dependence has recently been completed. Preliminary results suggest that the patients receiving the access to A-CHESS in the treatment group demonstrated significantly reduced heavy drinking days (i.e., the number of days during which a patient's drinking in a 2-hour period exceeded, for men, 4 standard drinks and for women, 3 standard drinks) compared to those receiving treatment as usual in the control group across an 8-month intervention period and a 4-month subsequent follow-up period (Gustafson et al., 2012). In addition, A-CHESS received long lasting usage—about 80% of patients still used A-CHESS weekly in the end of 4 months after intervention, as compared to 65% of asthma teenagers and 35% colon cancer survivors using respective smartphone applications at the same time period (McTavish et al., 2012). Nevertheless, despite the encouraging outcomes, the full potential of A-CHESS may not yet have been realized.

As a smartphone application that is accessible at 24/7, A-CHESS has the potential to offer instant support to patients and their counselors at the moment they most need it. A lapse often happens suddenly, but

may be predictable (Collins et al., 1998; Högström Brandt, Thorburn, Hiltunen, & Borg, 1999; Witkiewitz & Masyn, 2008). Being able to predict that a patient may lapse would enable A-CHESS to offer proactive support to patients before a lapse actually occurs. However, at its inception, no such mechanism was available that would allow A-CHESS to know when to provide such proactive intervention. Thus, the task undertaken in this present study was to evaluate patients' data collected via A-CHESS and develop a prediction function to be implemented in A-CHESS.

The ongoing data collection (both self-reported and system generated) being done with A-CHESS offered a great opportunity to develop such an intelligent feedback function. One important source of data was the Weekly Check-in. Patients normally completed this brief survey instrument in A-CHESS, called Weekly Check-in, once every 7 days on their smartphones. A-CHESS then transmitted the collected data to a remote, secure server via the Internet.

The questions used in the Weekly Check-in were adopted from a multi-dimensional instrument, the brief addiction monitor (BAM), developed in 2008 (Cacciola, Alterman, Oslin, & McKay, 2008). Two BAM items related to the status of drinking or taking drugs were combined into a single item in the Weekly Check-in to assess patients' substance use activities during the week. This substance use item was used as the indicator of patient lapse status. Five other BAM items (i.e., sleeping problems, depression, urge, risky situation, and relationship troubles) were conceptualized as risk related items. Five additional BAM items (i.e., confidence, AA meeting attendance, religious activities, other activities and time with family) were considered as protection related items (Cacciola et al., 2008). Factors very similar to these 10 protective and risk items outlined above have been found in the literature to be strongly associated with substance use behavior (Baker, Piper, McCarthy, Majeskie, & Fiore, 2004; Hunter-Reel, McCrady, & Hildebrandt, 2009; Laffaye, McKellar, Ilgen, & Moos, 2008; Moos & Moos, 2007; Oslin, Cary, Slaymaker, Colleran, & Blow, 2009; Robinson, Krentzman, Webb, & Brower, 2011; Witkiewitz & Marlatt, 2004).

Because the data in the Weekly Check-in were provided by patients on a weekly basis, it is possible that the recovery progress status of a patient (based on the risk and protection items in the current Weekly Check-in) may be predictive of the substance use status reported by the same patient in the next Weekly Check-in. This paper introduces a predictive model that was developed using the data collected in the A-CHESS Weekly Check-in and reports results of its evaluation. While this study was not intended to validate a certain theoretical framework or set of predictors, those concepts were presented as a prelude to describing the construction of the predictive model.

2. Materials and methods

2.1. Participants

A total of 170 patients were recruited and randomized into the intervention group in a randomized controlled trial designed to test A-CHESS (McTavish et al., 2012). Only the patients in the intervention group were included in this predictive modeling study because they were the only group given access to A-CHESS in the trial. The patients were recruited from two residential treatment organizations—one in the midwestern and the other in the northeastern U.S.—from February, 2010 through November, 2011. Participants had to be at least 18 years old, meet the criteria for DSM-IV alcohol dependence upon entering treatment, and provide two backup contacts for follow-up. Patients were excluded if they had a history of suicidality, a significant developmental or cognitive impairment that would limit the ability to use A-CHESS, or vision problems. Self-report data from the pretest survey of these 170 patients, revealed a mean age of 38 years old, mostly Caucasian (81%), and mostly male (61%). About

42% had some college education or greater. About 20% were employed. More than half (62%) reported at pretest that they had abused drugs beyond alcohol (39.4% using cocaine along with other drugs and 22.4% using drugs other than cocaine) at some point, and 49% reported mental health problems.

2.2. Procedures

An onsite project coordinator at each clinic identified eligible patients from the clinic's administrative database. About 2 weeks before an eligible patient left residential treatment, the coordinator discussed the study with the patient, including procedures, benefits and risks of participating, and data to be collected. Patients were told that A-CHESS was an intervention geared around (but not limited to) alcohol addiction. If the patient was willing to take part, written informed consent was then obtained. The study was conducted according to the Declaration of Helsinki of 1975 and approved by the institutional review board at the University of Wisconsin-Madison.

Before leaving the residential facilities, patients in the intervention group received training on how to use A-CHESS and the smartphones with a mobile broadband connection provided as part of the trial. During the 8-month intervention period, patients were expected to submit the Weekly Check-in (i.e. an ongoing patient monitoring instrument) in A-CHESS once every 7 days. To complete a Weekly Check-in, patients turned on the Weekly Check-in page in A-CHESS and selected the answers for the questions. After patients clicked the "submit" button in the end of the survey, the Weekly Check-in immediately verified completeness and prompted patients to fill in unanswered items before the survey could be submitted. Once submitted, the Weekly Check-in data were sent to a secure server at the University of Wisconsin-Madison. Seven days following the prior Weekly Check-in submission, patients would receive a message via A-CHESS as a reminder for another Weekly Check-in submission. To be included in the trial, patients must agree to receive such a prompt. However, completion of the Weekly Check-in was not required to remain in the trial. Therefore, it is possible that substantially more than 7 days could elapse between submissions. More details about the randomized controlled trial can be found in prior publications (McTavish et al., 2012).

2.3. Weekly Check-in instrument

The Weekly Check-in contains an instrument presented across three screens (Fig. 1). On the first screen, patients were asked to report their substance use status in the last 7 days in a dichotomized form (i.e., yes/no). On the second screen, patients reported their perceived level of risk related items (i.e., sleeping problems, depression, urge, risky situation, and relationship troubles) in the past week on a scale from 0 (thumbs up at the green zone) to 7 (thumbs down at the red zone). On the third screen, they reported their perceived level of protection related items (i.e., confidence, AA meeting, religious activities, other activities, and time with family) in the past week on a scale from 0 (thumbs down at the red zone) to 7 (thumbs up at the green zone).

2.4. Measures

2.4.1. Lapse status

In this study, the substance use item from the first screen in the Weekly Check-in was used as the indicator of a patient's lapse status. A lapse was defined to have occurred when a patient indicated that he/she used alcohol or took drugs in the past week. "Lapse" only indicated whether a participant reported that they drank or took drugs in the prior week and did not provide information about the amount or the frequency of use during the lapse. Determining if
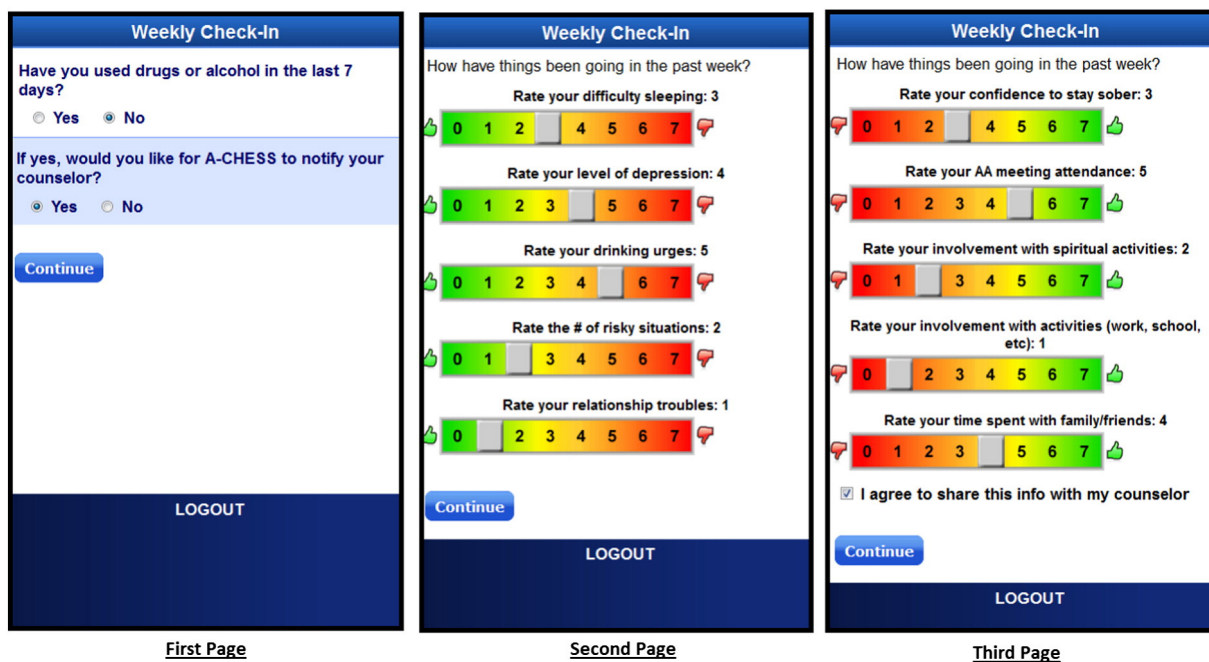
Fig. 1. Screen shots of the Weekly Check-in survey.

someone has relapsed involves much complexity and nuance that require data beyond that available in this present study.

2.4.2. Recovery progress

A score representing recovery progress was constructed by subtracting the summed score of the five risk items on the second screen of the Weekly Check-in (0–7 points each) from the summed score of the five protection items on the third screen of the Weekly Check-in (0–7 points each). The resulting recovery progress score can range from -35 to $+35$, with a higher value indicating better overall recovery progress.

2.4.3. Lapse history

Prior research has indicated that previous lapses were found to be a strong predictor of subsequent lapses (Högström Brandt et al., 1999; Witkiewitz, 2011). Furthermore, people with more frequent lapses were found to have a higher chance of subsequent lapses (Rounsaville, 2011; Witkiewitz, 2008). However, as patients stay sober, they may learn more effective coping and gain better self-efficacy to return back to abstinence (Witkiewitz & Marlatt, 2004). Based upon these aforementioned findings, a lapse history variable was constructed to represent patients' recent lapse history. The score was determined by the cumulative number of consecutive lapses reported at previous Weekly Check-ins. Each reported lapse would increment the lapse history by one. If a patient did not report a lapse in the most recent Weekly Check-in, his or her lapse history value was reduced to half of the previous value (e.g., from 4 to 2 or from 1 to .5). This variable was developed so that the dependency between reports from the same individuals across time can be taken into account in the model.

2.5. Conceptual model

The conceptual model used to construct the predictive function for weekly lapse prediction contains two of the measures described above, the recovery progress and the lapse history (Fig. 2). The dependent variable is the prediction of what lapse status will be reported in the next Weekly Check-in. Note that the one-item substance use question on the first screen of the Weekly Check-in asks patients about their substance use status in the last 7 days. If the

time between two consecutive reports was more than 14 days, the prediction was considered invalid because it was separated by too many days from the prior report. Therefore, only a Weekly Check-in with a subsequent Weekly Check-in within 14 days was considered an effective data point. If the next Weekly Check-in reported a lapse, it was taken as such; otherwise, it was considered a non-lapse.

2.6. Data preparation

A total of 3,219 Weekly Check-in reports submitted by 158 patients during their 8-month study period were retrieved from the A-CHESS server before August 17, 2011, the date on which the predictive model developed in this study was implemented in A-CHESS. Of these, 285 Weekly Check-in reports were excluded due to the lack of a subsequent report within 14 days. As a result, a total of 2,934 Weekly Check-in reports submitted by 152 patients between April 8, 2010 and August 17, 2011, were included in the following modeling process (Fig. 3).

2.7. Bayesian network modeling

The Bayesian network model has been recognized as one of the promising analytic methods in understanding often non-linear addiction behavior (Connor, Symons, Feeney, Young, & Wiles, 2007). Bayesian learning algorithms were developed to estimate the
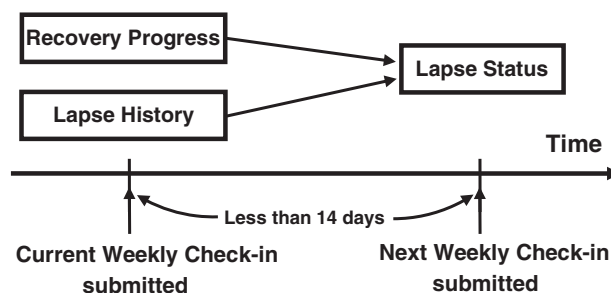

Fig. 2. Conceptual model of weekly lapse prediction.

```
┌─────────────────────────────────┐
│ Randomized into Intervention    │
│      Group (n=170)              │
└─────────────────────────────────┘
            │
            │      ┌─────────────────────────────────┐
            ├─────▶│ Did not submit any Weekly       │
            │      │ Check-in report (n=8)           │
            ▼      └─────────────────────────────────┘
┌─────────────────────────────────┐
│ Weekly Check-in reports within  │
│  the 8-month Intervention period│
│       (n=162, m=3,751)          │
└─────────────────────────────────┘
            │
            │      ┌─────────────────────────────────┐
            ├─────▶│ Weekly Check-in reports         │
            │      │ after A-CHESS system            │
            │      │ upgrade on August 17,           │
            ▼      │ 2011 (n=4, m=532)               │
┌─────────────────────────────────┐ └─────────────────────┘
│ Weekly Check-in reports before  │
│  A-CHESS system upgrade on      │
│     August 17, 2011             │
│     (n=158, m=3,219)            │
└─────────────────────────────────┘
            │
            │      ┌─────────────────────────────────┐
            ├─────▶│ Without a subsequent            │
            │      │ Weekly Check-in report in       │
            │      │ 14 days (n=6, m=285)            │
            ▼      └─────────────────────────────────┘
┌─────────────────────────────────┐
│ With a subsequent Weekly        │
│ Check-in report in 14 days      │
│      (n=152, m=2,934)           │
└─────────────────────────────────┘
```
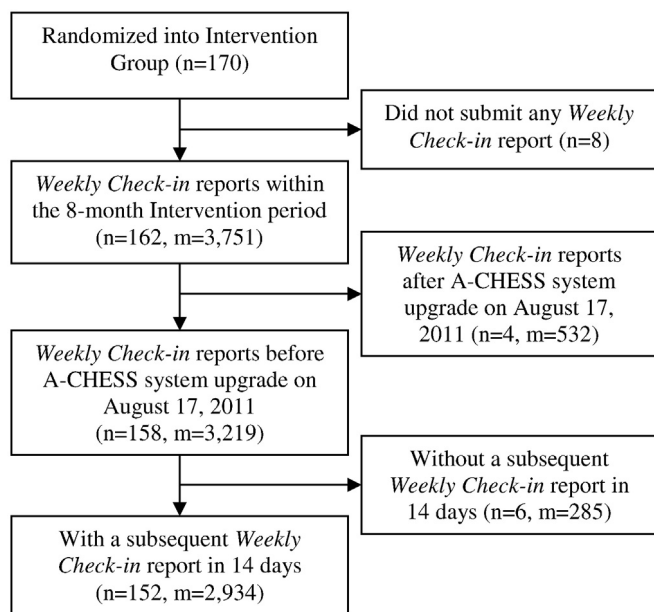
Fig. 3. Participants and Weekly Check-in reports flow chart. Note: n: the number of patients; m: the number of Weekly Check-in reports.

parameters (i.e., conditional probabilities) in a Bayesian network model (Koller & Friedman, 2010). Because no missing values were present in the Weekly Check-in data, we were able to use the fastest and least complex Bayesian learning algorithm, the "counting" algorithm, in this present study (Sayyad Amin, Alamdari, Mehranbod, Ayatollahi, & Nikooee, 2010; Spiegelhalter, Dawid, Lauritzen, & Cowell, 1993). Netica, a software package for building Bayesian network models that offers the "counting" algorithm, was used (Norsys Software Corp, 2010).

2.7.1. Model configuration

Variables in the Bayesian network model need to be measured as discrete categories rather than continuous variables. Initial estimations from our analyses of preliminary data suggested creating 10 categories for recovery progress and 3 categories for lapse history, as early tests of this model configuration provided good predictability (area under receiver operating characteristic curve [AUC] of 0.9). Therefore, the recovery progress score (-35 to $+35$) was divided into 10 equal discrete categories (e.g., -35 to -28, 0 to 7). Similarly, the lapse history score was divided into three discrete categories: values 0 to less than 0.5 were placed in the "stay sober" category; values 0.5 to less than 2 were placed in the "lapse once" category; and values 2 and more were placed in the "lapse more than once" category.

Prior studies have used 2 weeks of abstinence to mark the end of a prior drinking episode (Rounsaville, 2011; Stout, 2000; Zywiak et al., 2006). To model this, as mentioned earlier in the section 2.4.3., lapse history values changes based on patients' recent lapse status. Patients with a lapse history score of 1 (i.e., lapse once) reporting a non-lapse status in 2 consecutive Weekly Check-ins (corresponding to roughly 2 weeks) would have their lapse history score lowered to under 0.5, which moves it into the "stay sober" category. However, patients with an initial lapse may have a higher likelihood of lapsing again to avoid the feelings of guilt and failure from the initial lapse (Larimer, Palmer, & Marlatt, 1999). To model this possibility, patients reporting 2 or more subsequent lapse are moved from the "lapse once" category to the "lapse more than once" category. Reflecting potential effects of recent lapse history on the likelihood of a new lapse, patients in the "lapse more than once" category will take longer than 2 weeks to return to the "sober" category.

One advantage of Bayesian algorithms is the ability to integrate current data as well as prior information into the model (Uusitalo, 2007). In this present study, prior knowledge about the likelihood of a weekly lapse among study participants was unknown at the time when the model was being developed. Thus, a non-informative prior (50% lapse vs. 50% non-lapse) was used as a starting point. However, as new data were collected from the Weekly Check-in, they were incorporated into to the model.

2.7.2. Modeling process

The model was built using a two-stage process, using a model development stage followed by a model external validation stage. The 2,934 Weekly Check-in data were divided into two datasets based on the time when the reports were submitted: about 90% of the reports were used as the training dataset in the model development stage and 10% were held out as the testing dataset for the model external validation stage (Bleeker et al., 2003). In the model development stage, a stratified 10-fold cross-validation method was used to evaluate the model based on training data (Witten, Frank & Hall, 2011). Results from the 10-fold cross-validation were pooled together and analyzed in a receiver operating characteristic (ROC) curve analysis. The pooled area under the ROC curve (AUC) was used as the measure of model predictability (Bradley, 1997). With an AUC value of 1 indicating perfect prediction, an AUC of over 0.7 indicates that a model offers good prediction (Swets, 1988). After the 10-fold cross-validation, the final model was developed using all training data. In the model external validation stage, the final model developed earlier was tested using an ROC analysis with the testing data. The AUC and information about sensitivity and specificity were reported. Information about the implementation of the model and about an alerting feature in A-CHESS was provided in the Discussion section.

3. Results

A total of 2,934 Weekly Check-in reports submitted by 152 patients between April 8, 2010 and August 17, 2011 were included in this study. These 152 patients were demographically similar to the overall A-CHESS sample, with a mean age of 38 years old (ranging from 20 to 64), mostly Caucasians (83%), and mostly male (62%). About 41% of patients have received some college education or higher. Only 20% of patients were employed. About 63% of patients have abused drugs beyond alcohol (40.1% using cocaine along with other drugs and 23% using drugs other than cocaine) at some point. About 49% have mental health problems. Within the 8-month intervention period, out of these 152 patients, 23 (15%) submitted 32 or more Weekly Check-in reports; 38 (25%) submitted 24 to 31 reports; 30 (20%) submitted 16 to 23 reports; 31 (20%) submitted 8 to 15 reports; and 30 (20%) submitted 7 or fewer reports. Missing reports occurred more often later in the study. Overall, the lapse cases (i.e., a Weekly Check-in followed by a reported lapse in the following Weekly Check-in) showed a lower level of recovery progress with mean and standard deviation in parentheses, $-0.2 (12.2)$ versus $13.9 (12.2)$ and a higher level of lapse history, $2.9 (4.1)$ versus $0.1 (0.5)$, than the non-lapse cases.

3.1. Model development stage

The 2,640 Weekly Check-in reports (about 90% of 2,934 reports) that were submitted from April 8, 2010 through June 5, 2011 were used as the training dataset in the model development stage. About 6.9% (181/2,640) of them were followed by a lapse status reported in the next Weekly Check-in. These 181 lapse cases were submitted by 52 patients (mean = 3.48 and SD = 3.77). Based on the resulting data from the 10-fold cross-validation the pooled AUC (0.829) showed that the current model specification yielded good predictability. The

Table 1
Sensitivity and specificity at various thresholds.

Threshold	Sensitivity % of true lapses predicted	Specificity % of non-lapses predicted
>2%	100% (28/28)	38% (101/266)
>3%	96.4% (27/28)	59.4% (158/266)
>4-6%	75% (21/28)	88% (234/266)
>7-9%	71.4% (20/28)	93.6% (249/266)
>10%	71.4% (20/28)	94.4% (251/266)
>15%	71.4% (20/28)	95.5% (254/266)
>20%	71.4% (20/28)	97.4% (259/266)
>30%	71.4 (20/28)	98.1% (261/266)
>40%	67.9% (19/28)	98.9% (263/266)

10 AUCs generated in the 10-fold cross-validation ranged from 0.782 to 0.907, indicating that the model prediction was at an acceptable level given different combinations of the Weekly Check-in reports in the datasets. All data in the training dataset were then used to construct the Bayesian network model. In the model, recovery progress accounted for 2.91% of the variance in the lapse status, and lapse history accounted for 20.2%.

3.2. Model external validation stage

After completing the model development stage, an external validation was carried out on the resulting Bayesian network model. Test data for this stage consisted of 294 Weekly Check-in reports (about 10% of reports) submitted from June 5 to August 17, 2011. These reports contained a higher proportion of lapse cases (9.5%, 28/294) than the training data used in the model development stage. These 28 lapse cases were submitted by 10 patients (mean = 2.8 and SD = 3.01). The model testing result during this stage showed that the Bayesian network model developed earlier has good predictability with the AUC of 0.912 with 95% C.I. [0.862, 0.972], $p < .001$. The results of the ROC analysis carried out in the external validation stage were then used to determine a threshold value beyond which lapse would be predicted. The values of sensitivity and specificity were calculated for several threshold values based on the external validation results (Table 1).

Using a threshold of 5% means that when the likelihood of lapse, based on patients' current Weekly Check-in report, was over 5%, it was predicted that these patients would lapse in the following week. In Table 1, under a threshold of 5%, the model accurately predicted 21 out of 28 lapse cases (75% sensitivity) and 234 out of 266 non-lapse cases (88% specificity). The lower the threshold is held, the higher the sensitivity and the lower the specificity, and vice versa. Table 1 thus provides useful information to support decision making about choosing the thresholds and the associated tradeoff between sensitivity and specificity.

Based on the model, the probability of lapse for each combination of recovery progress and lapse history categories is provided in Table 2. For instance, a person with a recovery progress score between −21 and −14 as well as a lapse history lower than 0.5 has an 11% change of lapsing within the next week. In this way, Table 2 offers information about the characteristics of the lapse/non-lapse cases under a threshold value. For example, using the threshold of 5% described above, potential lapse cases were those with either a negative recovery progress value or a lapse history value of 0.5 or above.

4. Discussion

The intent of this project was to develop a model to estimate the chances of a patient lapsing within 1 week of the data collection. The model was to be used in a mobile health system (A-CHESS) to identify patients at high risk and then to take tailored action to reduce that risk. The resulting model appears to be stable and quite accurate. The ROC analysis results produced a table of sensitivity and specificity (Table 1) which was used to help select a threshold value beyond which A-CHESS would take action. The model is currently implemented along with an alerting feature as part of the A-CHESS system.

4.1. The implementation of the predictive model in A-CHESS

With good predictability of the lapse within the next week, the results of the model testing were encouraging enough to implement the Bayesian network model developed in this study within A-CHESS on August 17, 2011. Since that time, when a patient submitted a Weekly Check-in survey, A-CHESS estimated the likelihood that this patient would lapse in the coming week based on the recovery progress score from the report and the calculated lapse history. As these new cases were added, A-CHESS updated the model parameters for future prediction.

Based on the prediction, an alert feature was provided in A-CHESS. Using a pre-determined threshold (currently at 5%), if it was determined that an A-CHESS participant was likely to lapse, a text message was automatically sent to that participant about their risk of lapse. This text included suggestions of alternative activities in A-CHESS such as reaching out to others through the discussion group features, reviewing the information about ways to avoid lapse, planning alternative activities (tailored to their preferences such as

Table 2
The probability of lapse for all combinations of recovery progress and lapse history.

Recover progress [1]	Lapse history [2]	Probability	
		Lapse	Non-lapse
(−35 to −28)	Stay sober (0 to 0.5)	33%	67%
(−28 to −21)	Stay sober (0 to 0.5)	20%	80%
(−21 to −14)	Stay sober (0 to 0.5)	11%	89%
(−14 to −7)	Stay sober (0 to 0.5)	7%	93%
(−7 to 0)	Stay sober (0 to 0.5)	7%	93%
(0 to 7)	Stay sober (0 to 0.5)	4%	96%
(7 to 14)	Stay sober (0 to 0.5)	4%	96%
(14 to 21)	Stay sober (0 to 0.5)	2%	98%
(21 to 28)	Stay sober (0 to 0.5)	2%	98%
(28 to 35)	Stay sober (0 to 0.5)	2%	98%
(−35 to −28)	Lapse once (0.5 to 2)	25%	75%
(−28 to −21)	Lapse once (0.5 to 2)	50%	50%
(−21 to −14)	Lapse once (0.5 to 2)	33%	67%
(−14 to −7)	Lapse once (0.5 to 2)	41%	59%
(−7 to 0)	Lapse once (0.5 to 2)	38%	62%
(0 to 7)	Lapse once (0.5 to 2)	27%	73%
(7 to 14)	Lapse once (0.5 to 2)	18%	82%
(14 to 21)	Lapse once (0.5 to 2)	13%	87%
(21 to 28)	Lapse once (0.5 to 2)	8%	92%
(28 to 35)	Lapse once (0.5 to 2)	17%	83%
(−35 to −28)	Lapse more than once (>= 2)	80%	20%
(−28 to −21)	Lapse more than once (>= 2)	63%	38%
(−21 to −14)	Lapse more than once (>= 2)	80%	20%
(−14 to −7)	Lapse more than once (>= 2)	54%	46%
(−7 to 0)	Lapse more than once (>= 2)	67%	33%
(0 to 7)	Lapse more than once (>= 2)	59%	41%
(7 to 14)	Lapse more than once (>= 2)	60%	40%
(14 to 21)	Lapse more than once (>= 2)	50%	50%
(21 to 28)	Lapse more than once (>= 2)	33%	67%
(28 to 35)	Lapse more than once (>= 2)	50%	50%

1. Recovery progress (from −35 to +35) was divided into 10 equal discrete categories. 2. Lapse history were divided into three categories: values 0 to less than 0.5 were in the "sober" category; values 0.5 to less than 2 were in "lapse once" category; and values 2 and more were in "lapse more than once" category. 3. Probabilities in these two categories, (1) "recovery progress 14 to 21 and lapse more than once" and (2) "recovery progress 28 to 35 and lapse more than once", were based on the initial model specification (see section 2.7.1. Model configuration for details) because the training dataset does not contain cases in these two categories.

going for a run, going to the local alcoholics anonymous meeting), or reviewing their personal reasons for not using alcohol or drugs that they entered when they were trained and set-up on A-CHESS. In addition, the participant's counselor would receive an email alert about the participant's risk. The counselor could then reach out to the participant via email, text, or telephone. Lastly, a text alert was also sent to the CHESS study coordinator who would check in via text or email within one business day.

4.2. Comparisons, limitations, and future directions

Compared to the training data from older reports used to develop the model, the testing data used for external validation contains a larger proportion of lapses. This is not necessarily surprising as the test data used for external validation were derived from later reports, which represented data from patients who have been out of a rehabilitation facility longer and therefore may have a higher probability of lapses (Witkiewitz, 2011). Thus, the longer people use the A-CHESS, the opportunity for a predictive model to help actually increases. Beyond the implementation in A-CHESS, this prediction algorithm may help addiction counselors take a proactive approach with these patients, mitigating the expected trend of increasing potential lapse as patients' traverse their recovery journey.

This study does differ considerably with some aspects of the prior work. Many studies have focused on predicting relapse over a longer time window, often months or years (rather than lapses taking place over a short time period) (Kelly, Hoeppner, Urbanoski, & Slaymaker, 2011; Moos & Moos, 2007; Oslin et al., 2009; Pedersen & Hesse, 2009; Witkiewitz, 2011). The few studies that examined alcohol use over a short time (e.g., days or weeks) were primarily concerned with finding predictive factors and not providing an intervention (Collins et al., 1998; Kuntsche & Cooper, 2010). However, at least one study did attempt to build a predictive model. In this study, a group of researchers developed a logistic regression model to predict lapse episodes of 33 patients at a 3-day interval. They found patients with reported lapses had a higher chance of subsequent lapses, which is similar to the findings in the present study (Högström Brandt et al., 1999). However their model suffered from low sensitivity (Högström Brandt et al., 1999). In comparison, the model described in this present study reached a good level of sensitivity (over 70%) while maintaining a high level of specificity (over 90%).

Our results showed that under the current configuration (i.e., 10 categories of recovery progress and 3 categories of lapse history), the model provided good prediction of lapses within a week, which has achieved the goal of this study. As an exploratory effort, models based on alternative configurations (e.g., 14 categories of recovery progress) were developed and compared with the current model. The results showed that the current model either outperformed or was largely equivalent to the alternative models (results available upon request). Hence, the current model seems to have well-accepted, reliable predication.

While the data used in this study were collected weekly, more frequently collected data, such as using methods like ecological momentary assessment (EMA) (Collins et al., 1998; Shiffman, Stone, & Hufford, 2008), may provide more precise prediction. However, a methodology used to assess alcohol lapse risks on a weekly basis, such as that used in this present study, may match up well with out-patient treatment episodes (Wupperman et al., 2012) and weekend binge drinking patterns (Kuntsche & Cooper, 2010). Furthermore, compared to daily or even more frequent assessment methods, weekly assessment might prove to be less burdensome to patients and their counselors.

However, more frequent (e.g., daily or a couple of times in a day) and detailed (e.g. time of the substance use events) data collection could improve the predictive models. More frequent data collection may help to more accurately identify the moment of a lapse. More detailed data about substance use may help to provide a better determination of a lapse versus a relapse. Therefore, it is important

and could be very useful, in the future development of self-monitoring tools for alcohol addiction and other types of addiction, to consider a more frequent data collection period and collecting more detailed data about substance use. However, future development of such data collection systems should also study how to improve the survey burden caused by more frequent and intense data collection.

Other dynamic data currently available in A-CHESS may help to improve this model without increasing user's burden. Examples of such data include user generated text messages or discussion group postings and A-CHESS use data, such as how often patients interact with the application. Studies in cancer communication have shown that patients with higher eHealth system use and who have received and/or provided more emotional support in online discussion groups gained better health outcomes (Han et al., 2009; Namkoong et al., 2010). In further analysis, the likelihood of a reported lapse among the Weekly Check-in reports submitted irregularly (i.e. more than 14 days from prior reports) was found to be higher than that in the regular reports submitted within 14 days. When patients do not regularly submit the Weekly Check-in reports, they may not use A-CHESS regularly and/or may not receive the same level of the customized support and clinician reach-out from using the Weekly Check-in. Future studies should focus on exploring the potential to incorporate dynamic data in A-CHESS into the model in order to study more complex and dynamic relationships discussed here.

In addition to the above-mentioned need for more frequent and detailed data collection, several other limitations are noteworthy. First, patients in the A-CHESS trial were from only two treatment organizations. The generalizability of our model will need to be validated in future studies that involve patients from other clinics. Second, our dependent variable, the lapse status in the following week, was self-reported. Self-reported data may not provide the most accurate information about all lapse cases because some patients may be reticent to report a lapse. Even so, the self-reported lapses may still assist in predicting and preventing the chance of lapse among those who do accurately report. However, future studies that collect objective, clinical-based measures and the assessments from close families and friends may help to improve or validate the current results. Third, this study only included Weekly Check-in data as the predictor. Items in the Weekly Check-in may be treated as proximal factors that influence the substance use behavior at the moment (Donovan, 1996; Witkiewitz & Marlatt, 2004). Distal factors (e.g., addiction dependence) that may help to explain why a lapse occurs were not used in the present study. Future work to incorporate distal factors into the model building process should be explored for possible improvements in the predictive model. Last, the recovery progress factor is a composite score of the risk and protection items in the Weekly Check-in. Future study examining the factor structure and the relative importance of the risk or protection items, as compared to recovery progress score, may help further refine the model.

Despite these limitations, the predictive model developed in this study has moved us closer to the goal of providing lapse prediction to tailor interventions such as A-CHESS. The effort outlined here to unlock the power of data regularly collected on a smartphone application, use it to predict alcohol lapse, and implement the predictive model in the mobile application to guide intervention is unique. It is hoped that this serves as a starting point for further development. Information and mobile communication technology offers a great opportunity to develop novel ways to deliver addiction treatment to patients. With so much technology now available, we need to find better ways to assist people with the very difficult transition from alcohol addiction to sobriety.

had no further role in the study design; in the collection, analysis, and/or interpretation of the data; in the writing of the report; or in the decision to submit the paper for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIAAA or the National Institutes of Health. Preliminary results of this study have been presented as a poster in the 2011 Annual Symposium of American Medical Informatics Association.

References

Baker, T. B., Piper, M. E., McCarthy, D. E., Majeskie, M. R., & Fiore, M. C. (2004). Addiction motivation reformulated: An affective processing model of negative reinforcement. *Psychological Review*, 111, 33–51, http://dx.doi.org/10.1037/0033-295X.111.1.33.

Bleeker, S. E., Moll, H. A., Steyerberg, E. W., Donders, A. R. T., Derksen-Lubsen, G., Grobbee, D. E., & Moons, K. G. M. (2003). External validation is necessary in prediction research: A clinical example. *Journal of Clinical Epidemiology*, 56, 826–832, http://dx.doi.org/10.1016/S0895-4356(03)00207-5.

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30, 1145–1159, http://dx.doi.org/10.1016/S0031-3203(96)00142-2.

Cacciola, J. S., Alterman, A., Oslin, D., & McKay, J. (2008). *Brief alcoholism monitor (BAM) survey instrument*. Philadelphia, PA: The Treatment Research Institute, University of Pennsylvania.

Collins, R. L., Morsheimer, E. T., Shiffman, S., Paty, J. A., Gnys, M., & Papandonatos, G. D. (1998). Ecological momentary assessment in a behavioral drinking moderation training program. *Experimental and Clinical Psychopharmacology*, 6, 306–315, http://dx.doi.org/10.1037//1064-1297.6.3.306.

Connor, J. P., Symons, M., Feeney, G. F. X., Young, R. M., & Wiles, J. (2007). The application of machine learning techniques as an adjunct to clinical decision making in alcohol dependence treatment. *Substance Use & Misuse*, 42, 2193–2206, http://dx.doi.org/10.1080/10826080701658125.

Donovan, D. M. (1996). Assessment issues and domains in the prediction of relapse. *Addiction*, 91, S29–S36, http://dx.doi.org/10.1080/09652149638782.

Gustafson, D. H., Boyle, M. G., Shaw, B. R., Isham, A., McTavish, F., Richards, S., & Johnson, K. (2011a). An e-Health solution for people with alcohol problems. *Alcohol research & health: the journal of the National Institute on Alcohol Abuse and Alcoholism*, 33, 327–337 Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3536059.

Gustafson, D. H., McTavish, F. M., Atwood, A., Chih, M. -Y., Shah, D., Boyle, M., & Levy, M. (2012). Effects of a mHealth intervention for alcohol relapse prevention: A randomized trial. *2012 mHealth Summit*. Washington, DC: mHealth Summit.

Gustafson, D. H., Shaw, B. R., Isham, A., Baker, T., Boyle, M. G., & Levy, M. (2011b). Explicating an evidence-based, theoretically informed, mobile technology-based system to improve outcomes for people in recovery for alcohol dependence. *Substance Use & Misuse*, 46, 96–111, http://dx.doi.org/10.3109/10826084.2011.521413.

Han, J. Y., Hawkins, R. P., Shaw, B. R., Pingree, S., McTavish, F., & Gustafson, D. H. (2009). Unraveling uses and effects of an interactive health communication system. *Journal of Broadcasting & Electronic Media*, 53, 112–133, http://dx.doi.org/10.1080/08838150802643787.

Högström Brandt, A. M., Thorburn, D., Hiltunen, A. J., & Borg, S. (1999). Prediction of single episodes of drinking during the treatment of alcohol-dependent patients. *Alcohol (Fayetteville, N.Y.)*, 18, 35–42, http://dx.doi.org/10.1016/S0741-8329(98)00065-2.

Hunter-Reel, D., McCrady, B., & Hildebrandt, T. (2009). Emphasizing interpersonal factors: An extension of the Witkiewitz and Marlatt relapse model. *Addiction (Abingdon, England)*, 104, 1281–1290, http://dx.doi.org/10.1111/j.1360-0443.2009.02611.x.

Kelly, J. F., Hoeppner, B. B., Urbanoski, K. A., & Slaymaker, V. (2011). Predicting relapse among young adults: Psychometric validation of the Advanced WArning of RElapse (AWARE) scale. *Addictive Behaviors*, 36, 987–993, http://dx.doi.org/10.1016/j.addbeh.2011.05.013.

Koller, D., & Friedman, N. (2010). *Probabilistic graphical models: Principles and techniques*. Cambridge, MA: The MIT Press.

Kuntsche, E., & Cooper, M. L. (2010). Drinking to have fun and to get drunk: Motives as predictors of weekend drinking over and above usual drinking habits. *Drug and Alcohol Dependence*, 110, 259–262, http://dx.doi.org/10.1016/j.drugalcdep.2010.02.021.

Laffaye, C., McKellar, J. D., Ilgen, M. A., & Moos, R. H. (2008). Predictors of 4-year outcome of community residential treatment for patients with substance use disorders. *Addiction (Abingdon, England)*, 103, 671–680, http://dx.doi.org/10.1111/j.1360-0443.2008.02147.x.

Larimer, M. E., Palmer, R. S., & Marlatt, G. A. (1999). Relapse prevention. An overview of Marlatt's cognitive-behavioral model. *Alcohol Research & Health: the journal of the National Institute on Alcohol Abuse and Alcoholism*, 23, 151–160 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10890810.

Marlatt, G. A., & George, W. H. (1984). Relapse prevention: Introduction and overview of the model. *British Journal of Addiction*, 79, 261–273, http://dx.doi.org/10.1111/j.1360-0443.1984.tb00274.x.

McKay, J. R. (2005). Is there a case for extended interventions for alcohol and drug use disorders? *Addiction (Abingdon, England)*, 100, 1594–1610, http://dx.doi.org/10.1111/j.1360-0443.2005.01208.x.

McKay, J. R. (2009). Continuing care research: What we have learned and where we are going. *Journal of Substance Abuse Treatment*, 36, 131–145, http://dx.doi.org/10.1016/j.jsat.2008.10.004.

McTavish, F. M., Chih, M. -Y., Shah, D., & Gustafson, D. H. (2012). How patients recovering from alcoholism use a smartphone intervention. *Journal of Dual Diagnosis*, 8, 294–304, http://dx.doi.org/10.1080/15504263.2012.723312.

Moos, R. H., & Moos, B. S. (2007). Treated and untreated alcohol-use disorders: Course and predictors of remission and relapse. *Evaluation Review*, 31, 564–584, http://dx.doi.org/10.1177/0193841X07306749.

Namkoong, K., Shah, D. V., Han, J. Y., Kim, S. C., Yoo, W., Fan, D., & Gustafson, D. H. (2010). Expression and reception of treatment information in breast cancer support groups: How health self-efficacy moderates effects on emotional well-being. *Patient Education and Counseling*, 81(Suppl), S41–S47, http://dx.doi.org/10.1016/j.pec.2010.09.009.

Norsys Software Corp. (2010). Netica. Norsys Software Corp. Retrieved from https://www.norsys.com/.

Oslin, D. W., Cary, M., Slaymaker, V., Colleran, C., & Blow, F. C. (2009). Daily ratings measures of alcohol craving during an inpatient stay define subtypes of alcohol addiction that predict subsequent risk for resumption of drinking. *Drug and Alcohol Dependence*, 103, 131–136, http://dx.doi.org/10.1016/j.drugalcdep.2009.03.009.

Pedersen, M. U., & Hesse, M. (2009). A simple risk scoring system for prediction of relapse after inpatient alcohol treatment. *The American Journal on Addictions*, 18, 488–493, http://dx.doi.org/10.3109/10550490903205983.

Robinson, E. A. R., Krentzman, A. R., Webb, J. R., & Brower, K. J. (2011). Six-month changes in spirituality and religiousness in alcoholics predict drinking outcomes at nine months. *Journal of Studies on Alcohol and Drugs*, 72, 660–668 Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3125889/.

Rounsaville, D. (2011). *Lapse, relapse, and chasing the wagon: Post-treatment drinking and recovery (Doctoral Dissertation)*. University of Maryland. Retrieved from http://gradworks.umi.com/34/39/3439755.html

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68–78, http://dx.doi.org/10.1037/0003-066X.55.1.68.

Sayyad Amin, J., Alamdari, A., Mehranbod, N., Ayatollahi, S., & Nikooee, E. (2010). Prediction of asphaltene precipitation: Learning from data at different conditions. *Energy & Fuels*, 24, 4046–4053, http://dx.doi.org/10.1021/ef100106r.

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, 4, 1–32, http://dx.doi.org/10.1146/annurev.clinpsy.3.022806.091415.

Spiegelhalter, D. J. D. J., Dawid, A. P. P., Lauritzen, S. L. S. L., & Cowell, R. G. R. G. (1993). Bayesian analysis in expert systems. *Statistical Science*, 8, 219–247, http://dx.doi.org/10.1214/ss/1177010888.

Stout, R. L. (2000). What is a drinking episode? *Journal of Studies on Alcohol*, 61, 455–461 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10807219.

Substance Abuse and Mental Health Services Administration. (2011). *Substance Abuse and Mental Health Services Administration, Results from the 2010 National Survey on Drug Use and Health: Summary of National Findings. Substance Abuse* (Vol. I, p. NSDUH Series H–41, HHS Publication No. (SMA) 11–46). Rockville, MD: Substance Abuse and Mental Health Services Administration. http://dx.doi.org/10.1126/science.3287615.

Sullivan, L. E., Fiellin, D. A., & O'Connor, P. G. (2005). The prevalence and impact of alcohol problems in major depression: A systematic review. *The American Journal of Medicine*, 118, 330–341, http://dx.doi.org/10.1016/j.amjmed.2005.01.007.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. Science (New York, N.Y.), 240, 1285–1293 Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/3287615.

Uusitalo, L. (2007). Advantages and challenges of Bayesian networks in environmental modelling. *Ecological Modelling*, 203, 312–318, http://dx.doi.org/10.1016/j.ecolmodel.2006.11.033.

Witkiewitz, K. (2008). Lapses following alcohol treatment: Modeling the falls from the wagon. *Journal of Studies on Alcohol and Drugs*, 69, 594–604 Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2721009/.

Witkiewitz, K. (2011). Predictors of heavy drinking during and following treatment. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, 25, 426–438, http://dx.doi.org/10.1037/a0022889.

Witkiewitz, K., & Marlatt, G. A. (2004). Relapse prevention for alcohol and drug problems: That was Zen, this is Tao. *The American Psychologist*, 59, 224–235, http://dx.doi.org/10.1037/0003-066X.59.4.224.

Witkiewitz, K., & Masyn, K. E. (2008). Drinking trajectories following an initial lapse. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*, 22, 157–167, http://dx.doi.org/10.1037/0893-164X.22.2.157.

Witten, I. H., Frank, E., & Hall, M. A. (2011). Data Mining: Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems) (3rd ed). Burlington, MA: Morgan Kaufmann Publishers, 664.

Wupperman, P., Marlatt, G. A., Cunningham, A., Bowen, S., Berking, M., Mulvihill-Rivera, N., & Easton, C. (2012). Mindfulness and modification therapy for behavioral dysregulation: Results from a pilot study targeting alcohol use and aggression in women. *Journal of Clinical Psychology*, 68, 50–66, http://dx.doi.org/10.1002/jclp.20830.

Zywiak, W. H., Stout, R. L., Longabaugh, R., Dyck, I., Connors, G. J., & Maisto, S. A. (2006). Relapse-onset factors in Project MATCH: The relapse questionnaire. *Journal of Substance Abuse Treatment*, 31, 341–345, http://dx.doi.org/10.1016/j.jsat.2006.05.007.