Drug and Alcohol Dependence 264 (2024) 112431

Contents lists available at ScienceDirect

Drug and Alcohol Dependence

journal homepage: www.elsevier.com/locate/drugalcdep

Predicting cannabis use moderation among a sample of digital self-help subscribers: A machine learning study

Marleen I.A. Olthof[a,b,*], Lucas A. Ramos[b], Margriet W. van Laar[a], Anna E. Goudriaan[b,c,d], Matthijs Blankers[a,b,c]

[a] Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, the Netherlands
[b] Amsterdam UMC, Department of Psychiatry, University of Amsterdam, Amsterdam, the Netherlands
[c] Arkin Mental Health Care, Amsterdam, the Netherlands
[d] Amsterdam Public Health Research Institute, Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Cannabis
eHealth
Machine learning

ABSTRACT

Background: For individuals who wish to reduce their cannabis use without formal help, there are a variety of self-help tools available. Although some are proven to be effective in reducing cannabis use, effect sizes are typically small. More insight into predictors of successful reduction of use among individuals who frequently use cannabis and desire to reduce/quit could help identify factors that contribute to successful cannabis use moderation.
Methods: We analyzed data taken from a randomized controlled trial comparing the effectiveness of the digital cannabis intervention ICan to four online modules of educational information on cannabis. For the current study, we included 253 participants. Success was defined as reducing the grams of cannabis used in the past 7 days at baseline by at least 50 % at 6-month follow-up. To train and evaluate the machine learning models we used a nested k-fold cross-validation procedure.
Results: The results show that the two models applied had comparable low AUROC values of .61 (Random Forest) and .57 (Logistic Regression). Not identifying oneself as a cannabis user, not using tobacco products, high levels of depressive symptoms, high levels of psychological distress and high initial cannabis use values were the relatively most important predictors for success, although overall the associations were not strong.
Conclusions: Our study found only modest prediction accuracy when using machine learning models to predict success among individuals who use cannabis and desire to reduce/quit and show interest in digital self-help tools.

1. Introduction

There appears to be a treatment gap in Europe when it comes to seeking help for cannabis use problems. In 2022 an estimated 3.7 million European adults used cannabis (near) daily, while approximately 92,000 clients entered specialized treatment for cannabis-related problems (EMCDDA, 2024). And although it is evident that not all individuals who use cannabis (near) daily meet the criteria for cannabis use disorder (CUD) and require treatment, daily use forms a risk factor for CUD. The DSM-5 defines CUD as *'a problematic pattern of cannabis use leading to clinically significant impairment or distress, as manifested by at least two of the following, occurring within a 12-month period'* (American Psychiatric Association, 2013). The criteria, 11 in total, include using cannabis more frequently or in larger amounts than intended, symptoms of craving, symptoms of tolerance and withdrawal symptoms.

In the Netherlands a similar treatment gap is observed. Between 2007-2009 and 2019-2022, the last year prevalence of CUD has more than doubled in the Netherlands (ten Have et al., 2022). According to the most recent estimate of 2019–2022, the last year prevalence of CUD among adults was approximately 1.3 % (DSM-5 criteria) (ten Have et al., 2022). This translates to about 159,600 people. After an increase from 2001 to 2010, the numbers of people entering specialized addiction treatment for CUD has remained relatively stable in the past years (Wisselink et al., 2024). In 2022 there were 9231 individuals in treatment for CUD.

For individuals who wish to attempt to reduce their cannabis use

* Corresponding author at: Trimbos Institute, Netherlands Institute of Mental Health and Addiction, Utrecht, the Netherlands.
E-mail addresses: molthof@trimbos.nl (M.I.A. Olthof), lucas.ramos@nhlstenden.com (L.A. Ramos), mlaar@trimbos.nl (M.W. van Laar), a.e.goudriaan@amsterdamumc.nl (A.E. Goudriaan), mblankers@trimbos.nl (M. Blankers).

https://doi.org/10.1016/j.drugalcdep.2024.112431
Received 29 April 2024; Received in revised form 27 August 2024; Accepted 27 August 2024
Available online 5 September 2024

without formal help, there are a variety of self-help tools available. These tools range from books to digital interventions to self-help groups. And although some of these tools, such as digital interventions, have proven to be effective in reducing cannabis use, the effect sizes are typically small (Boumparis et al., 2019; Olthof et al., 2023). The digital interventions included in the meta-analysis performed by Boumparis et al. (2019) included interventions based on motivational interviewing (MI), cognitive behavioral therapy (CBT), personalized normative feedback (PNF), solution-focused therapy and parent-involvement programs. More insight into predictors of success among individuals who frequently use cannabis and desire to reduce or quit their cannabis use could help identify factors that contribute to successful outcomes.

Liebregts et al. (2015) conducted a qualitative longitudinal study in which they compared individuals who frequently used cannabis but successfully quit their cannabis use (desisters) to individuals who frequently used cannabis and had a persistent desire to use cannabis and unsuccessful quit attempts (persisters). The study sample consisted of 14 individuals selected from the broader Dutch Cannabis Dependence (CanDep) study on the course of frequent cannabis use and dependence, who participated in two additional in-depth interviews about their cannabis use. None of the participants were in treatment at the time of the study. The aim of the study was to better understand the underlying processes of desistance from cannabis use. Results showed that desisters experienced increasing control over their actions and set goals for themselves (Liebregts et al., 2015). In addition, they employed strategies that helped them to achieve their goals and were able to envision a different, non-using version of themselves. Their cannabis desistance was often influenced by life events for example traveling, disease of a loved one, or ending/beginning a relationship or job. Although persisters experienced similar life events, they lacked goals and strategies and struggled to envision a non-using version of themselves (Liebregts et al., 2015).

Rooke et al. (2011) conducted a quantitative study in which they compared characteristics of individuals who had successfully quit their cannabis use with those of current regular cannabis users who made at least one unsuccessful quit attempt. Successful quitters had a higher level of education, fewer symptoms of depression and stress, lower levels of cannabis dependence, lower day-to-day exposure to other individuals who use cannabis and higher exposure to formal treatment than the unsuccessful quitters (Rooke et al., 2011).

In addition, several studies have indicated characteristics that are predictive of success among individuals using digital self-help tools for problematic substance use, such as higher level of education, higher socioeconomic status, shared living situation, high interpersonal sensitivity and active participation in the online intervention's virtual community (Blankers et al., 2013; Bravin et al., 2015; Schwarzer and Satow, 2012). Riper et al. (2018) conducted an individual patient data meta-analysis of 19 randomized controlled trials to investigate moderators of treatment outcomes in internet-based interventions for problem drinking. Being male, less educated, and 55 years of age or older was associated with better treatment results (Riper et al., 2018).

Although previous studies (eg. Bravin et al., 2015; Riper et al., 2018; Schwarzer and Satow, 2012) provide valuable insights into potential predictors of success among individuals attempting to reduce their substance use, many of these previous studies have relied on traditional (multiple) regression models which were fitted and evaluated using the same dataset, instead of separate training and test data commonly used in machine learning approaches. Traditional (multiple) regression models are typically theory driven, where the model is constructed based on prior assumptions about the relationship between the variables in a given dataset (Ley et al., 2022). The model is usually fitted on all the available data, making it difficult to determine how generalizable the results are to other datasets. Machine learning models on the other hand are typically data-driven, where the data are split into a training and a test dataset (Coutanche and Hallion, 2020; Ley et al., 2022). The training dataset is used to optimize the model's fit through an iterative process. The test dataset is then used to evaluate the performance of the model on the other part of the data (Coutanche and Hallion, 2020; Ley et al., 2022). While in traditional (multiple) regression models the number of predictors is usually limited relative to the sample size and predictors cannot be highly correlated with other predictors, several machine learning models such as random forest models can evaluate the predictive power of a relatively large set of variables (Pearson et al., 2019). Therefore, machine learning may be able to address some of the shortcomings of traditional multiple regression approaches to predict which individuals who use cannabis are most likely to succeed in reducing their cannabis use. However, machine learning models also pose some challenges, including the need for larger sample sizes and occasional difficulty in interpreting the results due to their data-driven approach rather than being based on theoretical frameworks (Ley et al., 2022).

A previous study by Ramos et al. showed that machine learning models can be applied to predict success among individuals utilizing digital substance use self-help interventions, including a cannabis self-help intervention (Ramos et al., 2021). Success was defined as completion of all intervention modules and achieving the goal that the participant had made at the beginning of the intervention regarding cannabis use. Log data from the first three days of using the intervention were used to predict success. The best performing model had an AUC of 0.67, indicating limited added accuracy compared to chance level (i.e., .50) (Ramos et al., 2021; Šimundić, 2009). Ramos et al. did not include socio-demographic characteristics or other baseline characteristics, as these were unavailable for the majority of the participants. It is likely that the model accuracy could have benefitted from including socio-demographic data, because previous research has shown these to be predictive of outcome (Jonas et al., 2019).

To the best of our knowledge, only one study by Jonas et al. (2019) examined baseline characteristics predictive of success in an online cannabis intervention using a machine learning method. In total 31 predictors related to socio-demographics and substance use were included. Participants who (1) had the goal to abstain (vs. the goal to reduce), (2) had higher levels of self-reflection as measured by the Self Reflection and Insight Scale (vs. lower levels), (3) preferred a mild intoxicating effect (vs. a strong intoxicating effect) and (4) with high initial cannabis use at baseline (>22 days vs <22 days), had a higher probability of treatment success (Jonas et al., 2019). The model had an accuracy of 0.64, again indicating limited accuracy. Jonas et al. conclude that further potential predictors should be tested to enhance prediction accuracy.

In this study we aimed to use a wide variety of baseline characteristics with machine learning models to predict success among individuals who frequently use cannabis and desire to reduce or quit cannabis use. We also aimed to explore which (type of) characteristics are associated with success. We used data from a randomized controlled trial (RCT) in which the effectiveness of the online cannabis intervention ICan was tested compared to four online modules of educational information on cannabis (Olthof et al., 2023). The RCT found favourable 3-month effects of ICan in reducing grams of cannabis used, but no difference in use days compared to the control condition (Olthof et al., 2023). No differences were found between conditions at 6-month follow-up. This study will help to gain insight in whether a machine learning approach to prognostic modeling contributes to prediction accuracy overall in comparison to traditional multiple regression approaches. The study also aims to help gain insight into factors that contribute to successful outcomes among individuals who use cannabis and desire to reduce or quit their cannabis use and show interest in digital self-help tools. If successful, our findings could help inform the development or adaptation of these tools, to better meet the needs of individuals who frequently use cannabis and desire to reduce or quit.

2. Materials and methods

2.1. Population

We obtained self-reported data between December 2019 and November 2020. Follow-up measurements took place 6 weeks (treatment satisfaction only), 3 months and 6 months after randomization via online surveys. For the current study we only used the 6-month follow-up data, because this measurement point was also used for the primary outcome measure in the RCT (Olthof et al., 2023). Since the RCT found no significant difference in effectiveness between treatment conditions at 6-month follow-up, and to maintain an adequate sample size, the RCT data from the ICan intervention condition and control condition were combined for the current study. A total of 378 participants were available in the dataset, from which 253 were included in the study. Fig. 1 shows the flow diagram of the current study, including the numbers and reasons for exclusion of participants.

2.2. Interventions

The ICan intervention is a progressive web app based on motivational interviewing and cognitive behavioral therapy. ICan includes self-tests, goal-setting and exercises to develop a plan for achieving the goal. ICan also includes detailed information about seeking professional help for cannabis use problems. The educational information modules (control condition) consist of information about cannabis, its effects, and the risk of cannabis use disorder. The modules also provide general tips on reducing or quitting cannabis use. A detailed description of the trial and the interventions can be found elsewhere (Olthof et al., 2021). The trial received ethical approval from Medical Research Ethics Committees

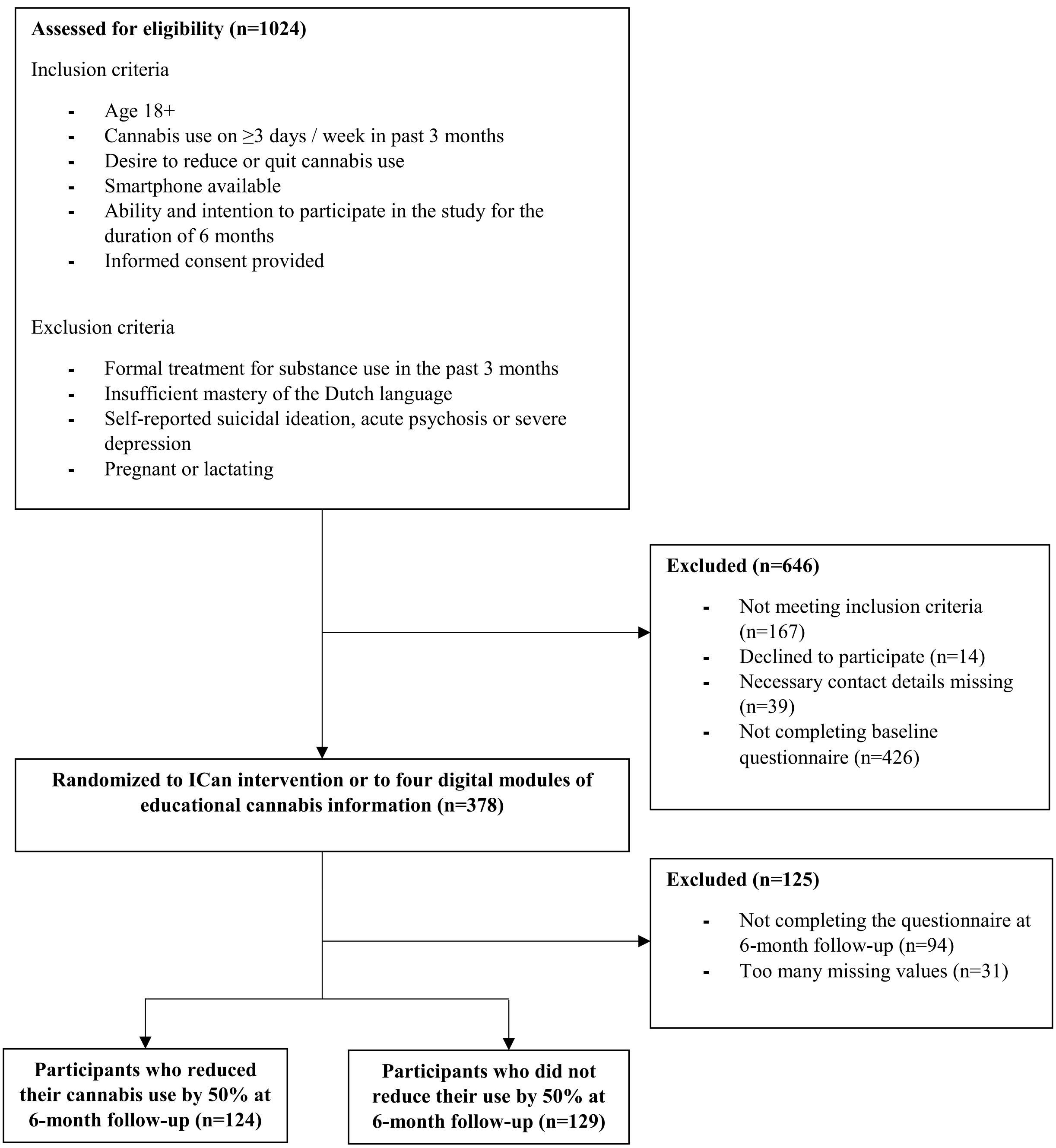

Fig. 1. Flow diagram.

United, NL67449.100.18.

2.3. Variables

Among others, the following baseline characteristics were collected prior to randomization:

Socio-demographics – age, gender and level of education.

Cannabis use variables – number of cannabis use days in the past 7 days assessed using the Timeline Follow-back method; number of cannabis use days in the past 30 days; average number of joints made from 1 g of cannabis; average number of joints smoked per use day; number of grams of cannabis used in the past 7 days; price paid per gram of cannabis product; amount spent in euros on cannabis per week; type of cannabis used (hashish or marijuana or both); route of administration.

Cannabis use related variables – cannabis use related problems assessed using the Cannabis Use Disorders Identification Test (CUDIT) (Adamson and Sellman, 2003); severity of dependence as measured by the self-reported number of DSM-5 criteria met for cannabis use disorder (American Psychiatric Association, 2013); the degree to which a person identifies oneself as a cannabis user assessed using the Cannabis Self-Concept Scale (CSCS) (Blevins et al., 2018), craving for cannabis as measured with the Marijuana Craving Questionnaire Short Form (MCQ-SF) (Heishman et al., 2009).

Attempts to reduce or quit cannabis use; past 3-month cannabis quit attempts lasting at least 24 hrs; number of days abstinent during most recent quit attempt; past 3-month attempts to reduce cannabis use lasting at least 24 hrs; number of joints smoked per week before most recent attempt to reduce; number of joints smoked per week during attempt to reduce; duration (days) of most recent attempt to reduce; self-efficacy to quit cannabis use assessed as in the study by Malmberg et al. (2012); self-reported withdrawal symptoms during most recent attempt to reduce or quit cannabis use.

Help-seeking related variables – attitudes towards seeking professional help for cannabis use related problems as measured with an adapted version of the Mental Help Seeking Attitudes Scale (MHSAS) (Hammer et al., 2018); attitudes, social norm, perceived behavioral control and intention to seek professional help for cannabis use related problems assessed using a self-constructed questionnaire developed according to Ajzen's theory of planned behavior.

Psychological distress and quality of life – psychological distress and symptoms of somatization, depression and anxiety were assessed using the BSI-18 (Derogatis, 2000).

Other substance use variables – self-reported use (number of use days in the past month) of tobacco, alcohol, amphetamine, ecstasy, hallucinogens, cocaine, crack, heroin, GHB, methadone, ketamine, nitrous oxide and unprescribed medication. In addition the AUDIT-C was included to measure hazardous drinking (Bush et al., 1998).

Other variables – The Marlowe-Crowne Social Desirability Scale (MCSDS) was included to be able to statistically control for social desirable responding in the randomized controlled trial (Crowne and Marlowe, 1960; Olthof et al., 2021). Knowledge about CBT principles and treatment options were measured with a self-constructed questionnaire.

2.4. Definition of success

Success was defined as a binary outcome. Based on the inspection of data, participants were considered to be successful if they reduced the amount (grams) of cannabis they used in the past 7 days at baseline by at least 50 % at 6-month follow-up. We selected this definition of success because it resulted in approximately equal-sized groups of participants who were classified as successful and unsuccessful and because a 50 % reduction in use can be deemed a clinically relevant result. We defined success based on cannabis consumption rather than cannabis use related problems, because the primary goal of the ICan intervention is to reduce cannabis use among participants regardless of whether they currently experience cannabis use related problems.

2.5. Machine learning models

Based on previous research and experience we applied two machine learning models (Ramos et al., 2021). First, a Logistic Regression model which has a linear approach, i.e. the model assumes a linear relationship between the predictor variables and the output variable, and is less robust when it comes to a large number of variables being included in the model. Second a Random forest model which tends to be more robust when dealing with a large number of variables and is able to identify non-linear relationships among the variables.

2.6. Statistical analysis

Due to the limited sample size the data was not analyzed separately by treatment condition. However, treatment condition was included as a potential predictor in the models. In addition, we included all the previously described baseline variables in the models. To train and evaluate the models we used a nested k-fold cross-validation procedure as described in the study by Ramos et al. (2021). For both the Logistic regression method and the Random forest method 10 models were optimized and tested using this cross-validation procedure. The average across all cross-validation iterations and 95 % Confidence Intervals are reported in Table 2 for the following five performance measurements:

- Sensitivity: proportion of truly successful participants that are correctly identified as successful by the model, this is also called the true positive rate.
- Specificity: proportion of truly unsuccessful participants that are correctly identified as unsuccessful by the model, this is also called the true negative rate.
- Negative predictive value (NPV): proportion of predicted unsuccessful participants that are truly unsuccessful.
- Positive predictive value (PPV): proportion of predicted successful participants that are truly successful.
- The Area Under the Receiver Operating Characteristics curve (AUROC): this is a more global measure of the model's prediction accuracy. The AUROC is calculated by plotting the sensitivity against 1-specificity for a number of classification thresholds (Mandrekar, 2010). AUROC values range from 0 to 1, a value of 0.5 indicates that the model does not predict better than random guessing, a value of 1.0 indicates perfect prediction.

We used Shapley Additive Explanations (SHAP) to visualize and interpret the output of the machine learning models. Shapley values were first used in the context of cooperative game theory to assign pay outs to players based on their contribution to the total pay-out (Molnar, 2019). In machine learning different features collaborate to predict a certain value. While most explanations created with Shapley values require to use all the features, Lundberg and Lee introduced a method (SHAP) that can provide explanations based on a selection of features (Lundberg and Lee, 2017; Molnar, 2019). The SHAP plot shows how a high or low value on a certain feature impacts and contributes to the prediction of -in our case- success.

3. Results

3.1. Participants

The last column of Table 1 shows the baseline demographic and clinical characteristics for the total sample. The age of the participants ranged between 18 and 64 years (M = 27.5, SD = 8.47), 69.6 % of them were male.

Table 1
Baseline demographic and clinical characteristics for each group (unsuccessful and successful) and total.

Characteristics	Unsuccessful (n=129)	Successful (n=124)	Total (n=253)
Male	92 (71.3 %)	84 (67.7 %)	176 (69.6 %)
Age, years (M ± SD)	26.93 ± 7.23	28.02 ± 9.59	27.46 ± 8.47
Level of education, n			
Low	24 (18.6 %)	24 (19.4 %)	48 (19.0 %)
Medium	50 (38.8 %)	49 (39.5 %)	101 (39.9 %)
High	55 (42.6 %)	51 (41.1 %)	104 (41.1 %)
Cannabis use frequency, days past week (M ± SD)	5.84 ± 1.61	6.26 ± 1.15	6.05 ± 1.41
Cannabis use quantity, grams past week (M ± SD)	4.77 ± 5.33	5.91 ± 5.68	5.33 ± 5.52
BSI Global severity Index Score (GSI)	26.79 ± 7.34	29.19 ± 7.80	27.96 ± 7.65
DSM-5, self-reported symptoms CUD (M ± SD)	5.97 ± 2.61	6.15 ± 2.62	6.06 ± 2.61

3.2. Success rates

The 6-month follow-up questionnaire was completed by 284 (75 %) of the participants, of whom 31 participants had too many missing values and were excluded. Of the remaining 253 participants, 124 participants successfully reduced their cannabis use (grams used in past 7 days) by 50 % or more, 129 participants were unsuccessful in reducing their cannabis use by 50 % or more. The first columns of Table 1 show the baseline demographic and clinical characteristics for each outcome group (unsuccessful & successful).

3.3. Model accuracy

Overall, the model accuracy measures indicated at best a modest fit of the models to the data. The values of all evaluation measures (AUROC, sensitivity, specificity, PPV and NPV) were slightly higher for the model using the Random Forest method than for the model using the Logistic Regression method, although the 95 % CI for the two models overlapped. The AUROC value of the Random Forest model was 0.61 (95 % CI 0.56–0.67). Table 2 shows the average values of the model evaluation measures and 95 % CI.

3.4. Feature importance

Fig. 2 shows the SHAP for the Random Forest model based on the 6-month follow-up data. The vertical axis shows the 20 most important features in the dataset, ranked with the most important features at the top. The horizontal axis shows the SHAP values. Features with a positive SHAP value are indicative of success (in our case: reducing cannabis use by ≥50 %), whereas features with a negative SHAP value are indicative of non-success (not reducing cannabis use by ≥50 %). Fig. 3 shows the relative impact of each feature on the model. In the interpretation we should take in account the overall modest predictive accuracy of the machine learning model.

At the top of the list with the most predictive features is the Cannabis Self Concept Scale (CSCS) score. High CSCS scores have negative SHAP values. This means that strongly identifying oneself as a cannabis user is associated with somewhat lower probabilities of reducing cannabis use by ≥50 % (non-success). The BSI-18 depression subscale score, and the Global Severity Index (GSI) are also relatively important predictors; high BSI/GSI scores, indicating higher levels of depressive symptoms and psychological distress, are associated with higher probabilities of reducing cannabis use by ≥50 % (success).

Use of tobacco products was also in the top 10 of predictive features. A higher number of days on which tobacco products were used in the past 30 days (not in combination with cannabis) was associated with non-success.

High initial cannabis use values – both the amount of grams of cannabis used in the week preceding baseline assessment and the number of joints smoked on the Sunday preceding baseline assessment - were associated with success. In addition, the route of administration as food/beverage was associated with success.

High scores on positive attitudes towards seeking help for cannabis use related problems and high scores on intention to seek help were associated with success. On the contrary, high scores on perceived behavioral control to seek help were associated with non-success.

4. Discussion

In this study we evaluated the use of machine learning approaches to prognostic modeling. We aimed to use a wide variety of baseline characteristics in machine learning models to predict success among individuals who use cannabis and desire to reduce or quit cannabis use and with interest digital cannabis self-help tools. Participants were considered to be successful if they reduced the amount (grams) of cannabis they used in the past 7 days at baseline by at least 50 % at 6-month follow-up. The results show that the two models applied had comparable AUROC values of .61 (Random Forest) and .57 (Logistic Regression). This indicates poor to sufficient diagnostic accuracy according to the benchmarks presented in Šimundić (2009). Hence, we can conclude that for our application, machine learning techniques have not contributed to a particularly accurate model to predict successful cannabis moderation. Although the current study only yielded modest prediction accuracy, these findings do not seem far out of line with findings in previous studies. Jonas et. al reported an accuracy of .64 in their study in which they used a machine learning inspired method to examine 31 potential baseline predictors of treatment success in an online cannabis intervention (Jonas et al., 2019). Ramos et al. reported an AUROC value of 0.67 in their machine learning study in which they used log data from the first 3 days of intervention use to predict success in a digital cannabis intervention (Ramos et al., 2021). While Jonas et al. (2019) used a sample size comparable to that in our study, Ramos et al. (2021) used a much larger sample size.

Despite the modest prediction accuracy, the models enabled us to explore the baseline characteristics that were relatively most strongly associated with success. The cannabis self-concept scale score was at the top of the list with most predictive features, meaning that strongly identifying oneself as cannabis user is associated with a lower probability of success in online self-help tools. This accords with previous research that showed that high cannabis self-concept scale scores are negatively associated with motivation to change cannabis use (Blevins et al., 2018). The results are also in line with the findings from the qualitative study by Liebregts et al. (2015) that demonstrated that unsuccessful cannabis quitters struggle to envision a non-using version of themselves.

Table 2
Average values of the evaluation measures and 95 % CI.

Time	Method	AUROC	Sensitivity	Specificity	PPV	NPV
6 mths	**RF**	**.61 (.56–.67)**	**.58 (.50–.66)**	**.62 (.54–.70)**	**.60 (.54–.65)**	**.61 (.55–.67)**
6 mths	LR	.57 (.49–.64)	.56 (.49–.62)	.60 (.49–.70)	.57 (.50–.64)	.58 (.51–.65)

Note: AUROC, Area Under the Receiver Operating Characteristics Curve; RF, Random Forest; LR, Logistic Regression; PPV, Positive Predictive Value; NPV, Negative Predictive Value.

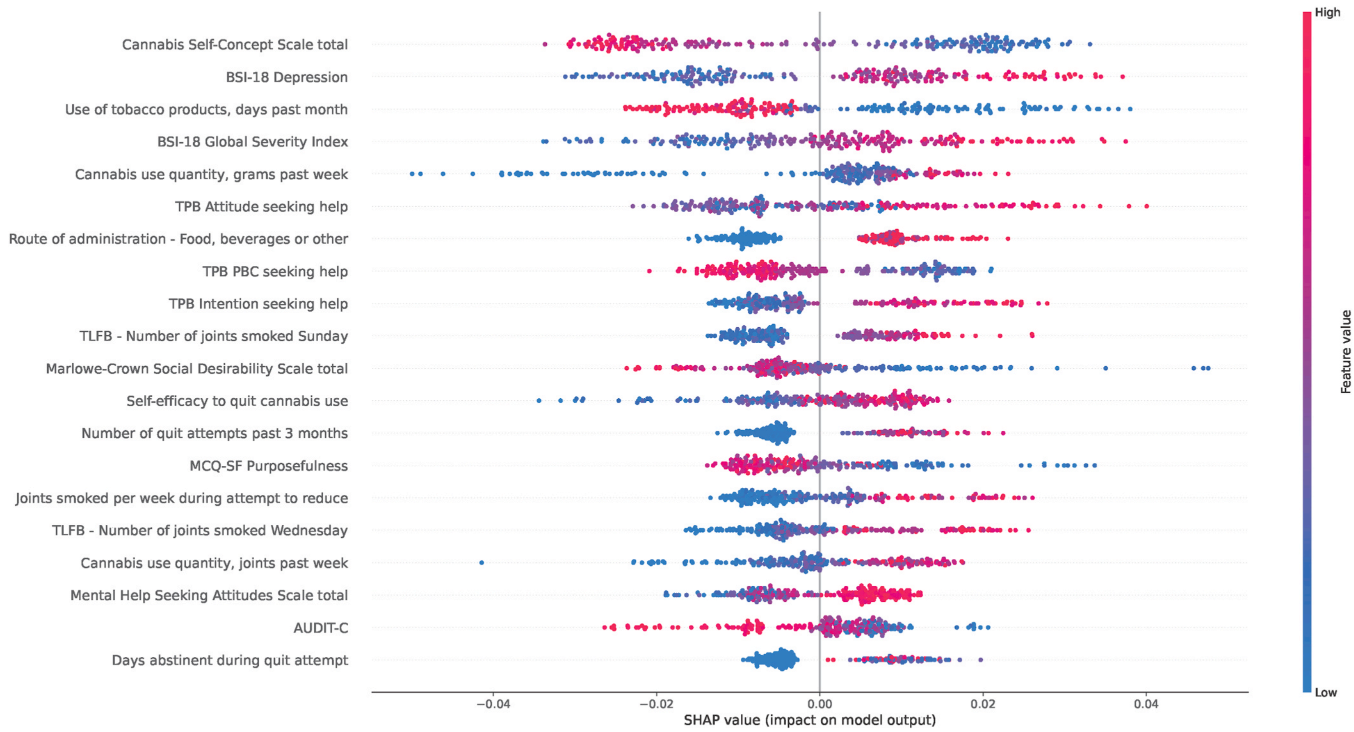

Fig. 2. SHAP feature importance for the Random Forest model based on the 6 month follow-up data. AUDIT-C, Alcohol Use Disorders Identification Test; BSI, Brief Symptom Inventory; MCQ-SF, Marijuana Craving Questionnaire Short Form; PBC, Perceived Behavioral Control; TLFB, Timeline Follow-Back; TPB, Theory of Planned Behavior.

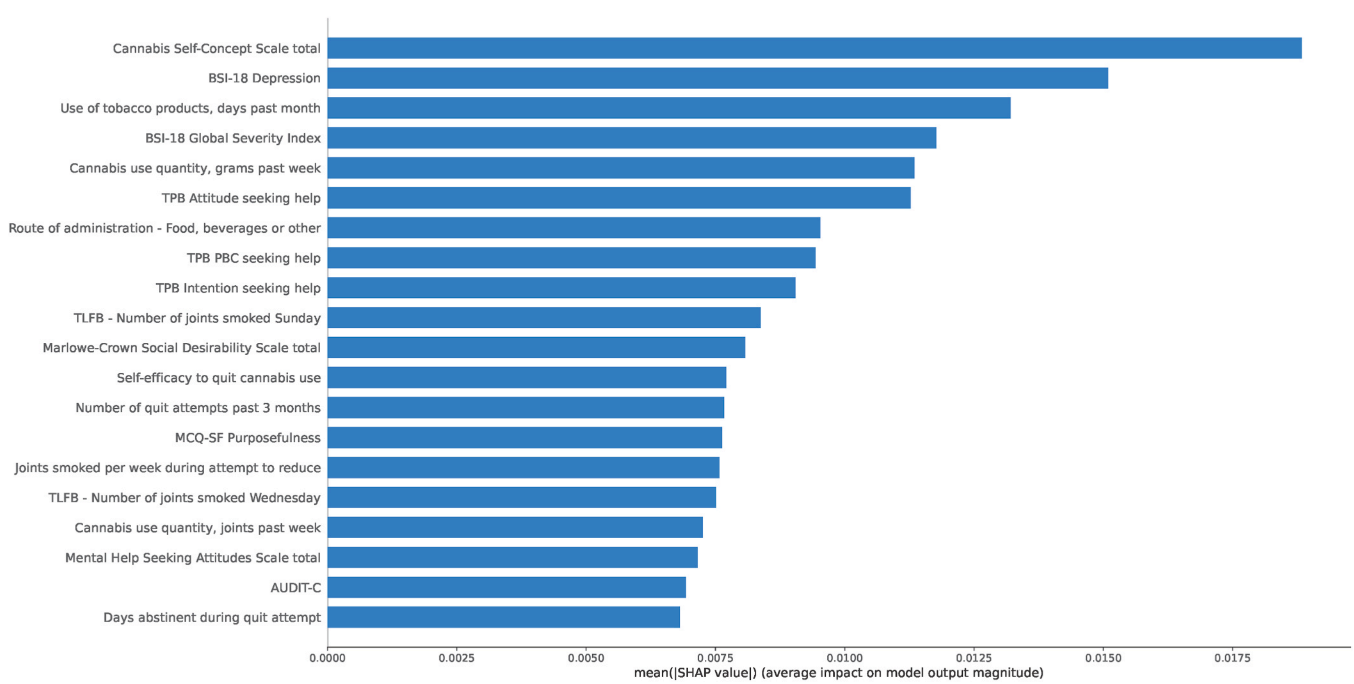

Fig. 3. Mean SHAP values, average impact of each feature on the model. AUDIT-C, Alcohol Use Disorders Identification Test; BSI, Brief Symptom Inventory; MCQ-SF, Marijuana Craving Questionnaire Short Form; Perceived Behavioral Control; TLFB, Timeline Follow-Back; TPB, Theory of Planned Behavior.

The use of tobacco products was also associated with lower probabilities of success in online self-help tools. These results corroborate the findings of previous reviews that show that tobacco use is negatively associated with cannabis cessation outcomes (McClure et al., 2020; Peters et al., 2012). In addition, a systematic review by Walsh et al. (2020) suggests that multi-substance interventions might be more effective than cannabis-only targeted interventions on cannabis cessation. These results suggest that adding components to the self-help tools that target tobacco use and the identity as a cannabis user may enhance the effectiveness.

High levels of depressive symptoms and psychological distress were associated with success (i.e. reducing cannabis use by $\geq$50 %) in online self-help tools. These findings are somewhat surprising given the fact that other research has shown that depressive symptoms are associated with unsuccessful quit attempts, poorer treatment response and increased risk of relapse in cannabis use disorder (Flórez-Salamanca et al., 2013; Rooke et al., 2011; Tomko et al., 2020; White et al., 2004). However, research has demonstrated that impaired mental health is one of the most important reasons for individuals to seek treatment for their cannabis use disorder (Van Der Pol et al., 2013). In this light, the presence of high levels of depression and psychological distress may reflect an increased sense of urgency and motivation in individuals who use cannabis to reduce or stop their cannabis use.

4.1. Limitations

As with any study, there are several limitations to our study that need to be addressed. First, the study has a relatively small sample size, particularly for a machine learning study where the data needs to be split into a training and a validation dataset. Second, the RCT data from the ICan intervention and the control condition were combined for the current study. The mechanisms driving success could differ between the two conditions, which could lead to different predictors of success. Third, we did not include any log data from the self-help interventions in the models, despite research showing its predictive value (Ramos et al., 2021). We made the decision to focus solely on baseline characteristics with the hope to identify a set of criteria that could potentially be used to screen for individuals with a high probability of benefiting from digital cannabis self-help tools. Furthermore, it was not possible to include log data in the analyses because in the control condition no patient-level log data was stored. Future studies that combine log data with baseline characteristics are recommended to enhance prediction accuracy, moreover future studies should use larger sample sizes and thus larger datasets.

4.2. Implications

Although machine learning techniques are promising, the current study demonstrates that they do not automatically lead to better prediction models. Moreover, it remains the question whether these techniques will help us gain more insight into predictors of successful behavioral change in substance use. It is evident that further research, using larger samples, and which have collected data on other factors that may be related to successful cannabis use moderation is required to determine whether machine learning techniques provide added value for this purpose. Future work may elucidate in which context and under what conditions the use of machine learning techniques may advance prognostic modeling of substance use behavior change outcomes.

5. Conclusion

All in all our study shows that using machine learning models to predict success (i.e. reducing cannabis use by ≥50 %) among individuals who use cannabis and desire to reduce or quit and with interest in digital self-help tools resulted in modest overall prediction accuracy, similar to previous studies.

Funding

The ICan study is funded by the Dutch Ministry of Health, Welfare and Sport. The funder had no role in the design, data collection, analysis and writing of this manuscript.

CRediT authorship contribution statement

Anna E. Goudriaan: Writing – review & editing, Supervision, Funding acquisition, Conceptualization, Methodology, Validation. **Margriet W. van Laar:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization, Methodology, Validation. **Matthijs Blankers:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization, Methodology, Validation. **Marleen I.A. Olthof:** Writing – original draft, Visualization, Validation, Investigation, Conceptualization, Methodology, Formal analysis. **Lucas A. Ramos:** Writing – review & editing, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no competing interests.

References

Adamson, S.J., Sellman, J.D., 2003. A prototype screening instrument for cannabis use disorder: the Cannabis Use Disorders Identification Test (CUDIT) in an alcohol-dependent clinical sample. Drug Alcohol Rev. 22, 309–315. https://doi.org/10.1080/0959523031000154454.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, fifth ed. American Psychiatric Publishing, Arlington, VA.

Blankers, M., Koeter, M.W.J., Schippers, G.M., 2013. Baseline predictors of treatment outcome in Internet-based alcohol interventions: a recursive partitioning analysis alongside a randomized trial. BMC Public Health 13. https://doi.org/10.1186/1471-2458-13-455.

Blevins, C.E., Abrantes, A.M., Anderson, B.J., Caviness, C.M., Herman, D.S., Stein, M.D., 2018. Identity as a cannabis user is related to problematic patterns of consumption among emerging adults. Addict. Behav. 79, 138–143. https://doi.org/10.1016/j.addbeh.2017.12.021.

Boumparis, N., Loheide-Niesmann, L., Blankers, M., Ebert, D.D., Korf, D., Schaub, M.P., Spijkerman, R., Tait, R.J., Riper, H., 2019. Short- and long-term effects of digital prevention and treatment interventions for cannabis use reduction: a systematic review and meta-analysis. Drug Alcohol Depend. 200, 82–94. https://doi.org/10.1016/j.drugalcdep.2019.03.016.

Bravin, J.I., Bunge, E.L., Evare, B., Wickham, R.E., Pérez-Stable, E.J., Muñoz, R.F., 2015. Socioeconomic predictors of smoking cessation in a worldwide online smoking cessation trial. Internet Interv. 2, 410–418. https://doi.org/10.1016/j.invent.2015.10.001.

Bush, K., Kivlahan, D.R., McDonell, M.B., Fihn, S.D., Bradley, K.A., 1998. The AUDIT Alcohol Consumption Questions (AUDIT-C). An effective brief screening test for problem drinking. Arch. Intern. Med. 158, 1789–1795.

Coutanche, M.N., Hallion, L.S., 2020. Machine learning for clinical psychology and clinical neuroscience. In: Wright, A.G.C., Hallquist, M.N. (Eds.), The Cambridge Handbook of Research Methods in Clinical Psychology. Cambridge University Press, Cambridge, pp. 467–482. https://doi.org/10.1017/9781316995808.041.

Crowne, D.P., Marlowe, D., 1960. A new scale of social desirability independent of psychopathology. J. Consult. Psychol. 24, 349–354. https://doi.org/10.1037/h0047358.

Derogatis, L.R., 2000. Brief Symptom Inventory-18 (BSI-18). https://doi.org/10.1037/t07502-000.

EMCDDA, 2024. European Drug Report 2024: Trends and Developments, European Monitoring Centre for Drugs and Drug Addiction.

Flórez-Salamanca, L., Secades-Villa, R., Budney, A.J., Garcia-Rodriguez, O., Wang, S., Blanco, C., 2013. Probability and predictors of cannabis use disorders relapse: results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). Drug Alcohol Depend. 132, 127–133. https://doi.org/10.1016/j.drugalcdep.2013.01.013.

Hammer, J.H., Parent, M.C., Spiker, D.A., 2018. Mental help seeking attitudes scale (MHSAS): development, reliability, validity, and comparison with the ATSPPH-SF and IASMHS-PO. J. Couns. Psychol. 65, 74–85. https://doi.org/10.1037/cou0000248.

Heishman, S.J., Evans, R.J., Singleton, E.G., Levin, K.H., Copersino, M.L., Gorelick, D.A., 2009. Reliability and validity of a short form of the Marijuana Craving Questionnaire. Drug Alcohol Depend. 102, 35–40. https://doi.org/10.1016/j.drugalcdep.2008.12.010.

Jonas, B., Tensil, M.D., Leuschner, F., Strüber, E., Tossmann, P., 2019. Predictors of treatment response in a web-based intervention for cannabis users. Internet Interv. 18, 100261 https://doi.org/10.1016/j.invent.2019.100261.

Ley, C., Martin, R.K., Pareek, A., Groll, A., Seil, R., Tischer, T., 2022. Machine learning and conventional statistics: making sense of the differences. Knee Surg. Sport. Traumatol. Arthrosc. 30, 753–757. https://doi.org/10.1007/s00167-022-06896-6.

Liebregts, N., van der Pol, P., de Graaf, R., van Laar, M., van den Brink, W., Korf, D.J., 2015. Persistence and desistance in heavy cannabis use: the role of identity, agency, and life events. J. Youth Stud. 18, 617–633. https://doi.org/10.1080/13676261.2014.992320.

Lundberg, S.M., Lee, S., 2017. A unified approach to interpreting model predictions. Adv. Neural Inf. Process. Syst. 30, 1–10.

Malmberg, M., Overbeek, G., Vermulst, A.A., Monshouwer, K., Vollebergh, W.A.M., Engels, R.C.M.E., 2012. The theory of planned behavior: precursors of marijuana use in early adolescence? Drug Alcohol Depend. 123, 22–28. https://doi.org/10.1016/j.drugalcdep.2011.10.011.

Mandrekar, J.N., 2010. Receiver operating characteristic curve in diagnostic test assessment. J. Thorac. Oncol. 5, 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d.

McClure, E.A., Rabin, R.A., Lee, D.C., Hindocha, C., 2020. Treatment implications associated with cannabis and tobacco co-use. Curr. Addict. Rep. 7, 533–544. https://doi.org/10.1007/s40429-020-00334-8.

Molnar, C., 2019. Interpretable Machine Learning. A Guide for Making Black Box Models Explainable. ⟨https://christophm.github.io/interpretable-ml-book/⟩

Olthof, M.I.A., Blankers, M., van Laar, M.W., Goudriaan, A.E., 2021. ICan, an internet-based intervention to reduce cannabis use: study protocol for a randomized controlled trial. Trials 22, 1–12. https://doi.org/10.1186/s13063-020-04962-3.

Olthof, M.I.A., Goudriaan, A.E., van Laar, M.W., Blankers, M., 2023. A guided digital intervention to reduce cannabis use: the ICan randomized controlled trial. Addiction 118, 1775–1786. https://doi.org/10.1111/add.16217.

Pearson, R., Pisner, D., Meyer, B., Shumake, J., Beevers, C.G., 2019. A machine learning ensemble to predict treatment outcomes following an Internet intervention for depression. Psychol. Med. 49, 2330–2341. https://doi.org/10.1017/S003329171800315X.

Peters, E.N., Budney, A.J., Carroll, K.M., 2012. Clinical correlates of co-occurring cannabis and tobacco use: a systematic review. Addiction 107, 1404–1417. https://doi.org/10.1111/j.1360-0443.2012.03843.x.

Ramos, L.A., Blankers, M., van Wingen, G., de Bruijn, T., Pauws, S.C., Goudriaan, A.E., 2021. Predicting success of a digital self-help intervention for alcohol and substance use with machine learning. Front. Psychol. 12, 1–11. https://doi.org/10.3389/fpsyg.2021.734633.

Riper, H., Hoogendoorn, A., Cuijpers, P., Karyotaki, E., Boumparis, N., Mira, A., Andersson, G., Berman, A.H., Bertholet, N., Bischof, G., Blankers, M., Boon, B., Boß, L., Brendryen, H., Cunningham, J., Ebert, D., Hansen, A., Hester, R., Khadjesari, Z., Kramer, J., Murray, E., Postel, M., Schulz, D., Sinadinovic, K., Suffoletto, B., Sundström, C., de Vries, H., Wallace, P., Wiers, R.W., Smit, J.H., 2018. Effectiveness and treatment moderators of internet interventions for adult problem drinking: an individual patient data meta-analysis of 19 randomised controlled trials. PLoS Med. 15, 1–26. https://doi.org/10.1371/journal.pmed.1002714.

Rooke, S.E., Norberg, M.M., Copeland, J., 2011. Successful and unsuccessful cannabis quitters: comparing group characteristics and quitting strategies. Subst. Abus. Treat. Prev. Policy 6, 30. https://doi.org/10.1186/1747-597X-6-30.

Schwarzer, R., Satow, L., 2012. Online intervention engagement predicts smoking cessation. Prev. Med. 55, 233–236. https://doi.org/10.1016/j.ypmed.2012.07.006.

Šimundić, A.M., 2009. Measures of diagnostic accuracy: basic definitions. EJIFCC 19, 203–211.

ten Have, M., Tuithof, M., van Dorsselaer, S., Schouten, F., De Graaf, R., 2022. NEMESIS. De psychische gezondheid van de Nederlandse bevolking. [Netherlands Mental Health Survey and Incidence Study]. Utrecht, the Netherlands.

Tomko, R.L., Baker, N.L., Hood, C.O., Gilmore, A.K., McClure, E.A., Squeglia, L.M., McRae-Clark, A.L., Sonne, S.C., Gray, K.M., 2020. Depressive symptoms and cannabis use in a placebo-controlled trial of N-Acetylcysteine for adult cannabis use disorder. Psychopharmacology 237, 479–490. https://doi.org/10.1007/s00213-019-05384-z.

Van Der Pol, P., Liebregts, N., de Graaf, R., Korf, D.J., van den Brink, W., van Laar, M., 2013. Facilitators and barriers in treatment seeking for cannabis dependence. Drug Alcohol Depend. 133, 776–780. https://doi.org/10.1016/j.drugalcdep.2013.08.011.

Walsh, H., Mcneill, A., Purssell, E., Duaso, M., 2020. A systematic review and Bayesian meta-analysis of interventions which target or assess co-use of tobacco and cannabis in single- or multi-substance interventions. Addiction 115, 1800–1814. https://doi.org/10.1111/add.14993.

White, A.M., Jordan, J.D., Schroeder, K.M., Acheson, S.K., Georgi, B.D., Sauls, G., Ellington, R.R., Swartzwelder, H.S., 2004. Predictors of relapse during treatment and treatment completion among marijuana-dependent adolescents in an intensive outpatient substance abuse program. Subst. Abus. 25, 53–59. https://doi.org/10.1300/J465v25n01.

Wisselink, D.J., Kuijpers, W.G.T., Kerssies, J.P., van der Slink, J.B., 2024. Kerncijfers Verslavingszorg 2017–2022. Landelijk Alcohol en Drugs Informatie Systeem (LADIS). Stichting Informatievoorziening Zorg, Houten, the Netherlands.